SPATIAL DISPLAYS AND SPATIAL INSTRUMENTS

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
WASHINGTON, DC

1989

**20100715227**

# Spatial Displays and Spatial Instruments

*Edited by*
Stephen R. Ellis
and
Mary K. Kaiser
*Ames Research Center*
*Moffett Field, California*

Arthur Grunwald
*Technion*
*Haifa, Israel*

**NASA**

National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California

1989

#·N90-22918-N90-2296,

NASA Conference Publication 10032

N90-22918

# Spatial Displays and Spatial Instruments

*Proceedings of a conference held*
*at Asilomar, California*
*August 31 - September 3, 1987*

NASA

# PREFACE

By the time this *Proceedings* is published, almost 2 years will have elapsed since the NASA-U.C. Berkeley Conference on Spatial Displays and Spatial Instruments held August 31-September 3, 1987, at the Asilomar Conference Center in Pacific Grove, California. The publication of the papers included in this proceedings will be a major step toward completion of a book to be based on material presented at the conference. Though the book itself will have a totally different organization, this *Proceedings* represents a kind of elaborate rough draft for it. The *Proceedings* are intended to provide not only the first comprehensive record of the conference, but also a written forum for the participants to provide corrections, updates, or short comments to be incorporated into the book's chapters.

I wish to sincerely thank again all the conference participants and especially Art Grunwald and Mary Kaiser, whose assistance and persistent reminders that the paper review must go forward have been helpful. Others who helped with the administrative details of the conference were Fidel Lam, Constance Ramos, Terri Bernaciak, and Michael Moultray. We also should thank the staff at Asilomar and the Ames Technical Information Division. I hope that the personal contacts and interchange of information initiated at the conference continues into the future and I look forward during the next 3 months to receiving addenda to be included in the book.

Stephen R. Ellis
Conference Organizer

Preceding page blank

# TABLE OF CONTENTS

# INTRODUCTION

# PICTORIAL COMMUNICATION: PICTURES AND THE SYNTHETIC UNIVERSE

Stephen R. Ellis
NASA Ames Research Center
Moffett Field, California
and
U. C. Berkeley School of Optometry
Berkeley, California

## SUMMARY

Principles for the design of dynamic spatial instruments for communicating quantitative information to viewers are considered through a brief review of the history of pictorial communication. Pictorial communication is seen to have two directions: 1) from the picture to the viewer and 2) from the viewer to the picture. Optimization of the design of interactive instruments using pictorial formats requires an understanding of the manipulative, perceptual, and cognitive limitations of human viewers.

## PICTURES

People have been interested in pictures for a long time (fig. 1). This interest has two related aspects. On one hand we have an interest in the picture of reality provided to us in bits and pieces by our visual and gross body orienting systems–and their technological enhancements. Indeed, Western science has provided us with ever clearer pictures of reality through the extension of our senses by specialized instruments.

On the other hand, we also have an interest in pictures for communication, pictures to transmit information among ourselves as well as between us and our increasingly sophisticated information-processing machines. This second aspect will be our prime focus, but some discussion of the first is unavoidable.

It is useful to have a working definition of what a picture is and I will propose the following: A picture is produced through establishment of a relation between one space and another so that some spatial properties of the first are preserved in the second, which is its image. A perspective projection is one of many ways this definition may be satisfied (fig. 2).

The definition may be fleshed out, as cartographers do, by exactly stating what properties are preserved, but the basic idea is that, though the defining relation of the layout of the picture may discard some of the original information, this relation is not arbitrary. The challenge in the design of a picture is the decision what to preserve and what to discard.

Artists, of course, have been making these decisions for thousands of years, and we can learn much from this history. One curious aspect of it, one that I certainly found strange when I learned

of it, is that early art was not focused on the preservation of spatial properties that I have asserted to be the essence of a picture.

As art historians have pointed out, early art was often iconographic, depicting symbols, as these Egyptian symbols for fractions illustrate, rather than aspiring to three-dimensional realism (fig. 3) (Gombrich, 1969). This early history underscores a second aspect of pictures which we must consider: their symbolic content. Because of the potentially arbitrary relation between a symbol and what it denotes, a symbol itself is not a picture. Symbols, nevertheless, have from the very beginning wormed their way into many pictures, and we now must live with both the symbolic and geometric aspects of pictorial communication. Furthermore, the existence of the symbolic content of the picture has the useful role of reminding the viewer of the essentially duplicitous nature of a picture since, though it inherently represents an alternative space, it itself is an object with a flat surface and fixed distance from the viewer.

The third basic element of pictorial communication is computational. The picture must be created. In the past the computation of a picture has primarily been a manual activity limited by the artist's manual dexterity, observational acumen, and pictorial imagination. The computation has two separable parts: 1) the shaping and placement of the components of the image, and 2) the rendering, that is, the coloring and shading of the parts (fig. 4).

While this second part is clearly important and can contribute in a major way to the success of a picture, it is not central to the discussion I wish to develop. Though the rendering of the image can help establish the virtual or illusory space that the picture depicts and can literally make the subject matter reach out of the picture plane, it is not the primary influence on the definition of this virtual space. Shaping and placement are. These elements reflect the underlying geometry used to create the image and determine how the image is to be rendered. By their manipulation artists can define—or confuse—the virtual space conveyed by their pictures.

While the original problems of shaping, positioning, and rendering still remain (figs. 5 and 6), the computation of contemporary pictures is no longer restricted to manual techniques. The introduction of computer technology has enormously expanded the artist's palette, and provided a new 3D canvas on which to create dynamic synthetic universes; yet the perceptual and cognitive limits of the viewers have remained much the same. Thus, there is now a special need for artists, graphic designers, and other creators of pictures for communication to understand these limitations of their viewers. Here is where the scientific interest in the picture of reality and the engineering interest in the picture for communication converge.

## SPATIAL INSTRUMENTS

In order to understand how the spatial information presented in pictures may be communicated, it is helpful to distinguish between images which may be described as spatial displays and those that were designed to be spatial instruments. One may think of a spatial display as any dynamic, synthetic, systematic mapping of one space onto another. A picture or a photograph is a spatial display of an instant of time (fig. 7). A silhouette cast by the sun is not, because it is a natural phenomenon not synthesized by humans.

A spatial instrument, in contrast, is a spatial display that has been enhanced either by geometric, symbolic, or computational techniques to ensure that the communicative intent of instrument is realized. A simple example of a spatial instrument is an analog clock (fig. 8). In a clock the angular positions of the arms are made proportional to time, and the viewer's angle-estimation task is assisted by radial tic marks designating the hours and minutes.

A second aspect of the definition of a spatial instrument, which the clock example also illustrates, is that the communicated variable—time—is made proportional to a spatial property of the display, such as an angle, areas, or length and is not simply encoded as a character string.

The spatial instruments on which we wish to focus attention are generally interactive. That is to say, the communicated information flows both to and fro between the viewer and the instrument. Some of this bidirectional flow exists for practically all spatial instruments, since movement of the viewer can have a major impact on the appearance of the display. However, the displays I wish to consider are those incorporating at least one controlled element, such as a cursor, which is used to extract information from and input information to the instrument.

Spatial instruments have a long history. One of the first ever made, dating from 60-80 BC, was an astrolabe-like device uncovered in 1901 near Antikythera, Greece. However, it was not fully described until the late '50's by De Solla Price (1959), who was able to deduce much of its principles of operation by x-raying the highly corroded remains (fig. 9). Here the communicated variables were the positions of heavenly bodies. Nothing approaching the complexity of this device is known until the 16th Century. It represents a highly sophisticated technology otherwise unknown in the historical record.

Though many subsequent spatial instruments have been mechanical and, like the Prague town hall clock (fig. 8), have similarly been associated with astronomical calculations (King, 1978), this association is not universal. Maps, when combined with mechanical aids for their use, certainly meet the definition of a spatial instrument (fig. 10). The map projection may be chosen depending upon the spatial property of importance. For example, straight-line mapping of compass courses (rhumb lines), which are curved on many maps, can be preserved in Mercator projections (Dickinson, 1979; Bunge, 1965). Choice of these projections illustrates a geometric enhancement of the map. The overlaying of latitude and longitude lines illustrates a symbolic enhancement (figs. 11-13). But more modern media may also be adapted to enhance the spatial information that they portray, as illustrated by the reference grid used by Muybridge in his photographs (Muybridge, 1975) (fig. 14).

Contemporary spatial instruments are found throughout the modern aircraft cockpit (fig. 15), the most notable probably being the attitude direction indicator which displays a variety of signals related to the aircraft's attitude and orientation. More recent versions of these standard cockpit instruments have been realized with CRT displays, which have generally been modeled after their electromechanical predecessors (Boeing, 1983). But future cockpits promise to look more like offices than anything else (fig. 16). In these offices the computer graphics and CRT display media, however, allow the conception of totally novel display formats for totally new, demanding aerospace applications.

For instance, a pictorial spatial instrument to assist informal, complex, orbital navigation in the vicinity of an orbiting spacecraft has been described (fig. 17) (see also Paper 37, Grunwald and Ellis, 1988). Other graphical visualization aids for docking and orbital maneuvering, as well as

other applications, have been demonstrated by Eyles (1986) (see also Paper 36). These new instruments can be enhanced in three different ways: geometric, symbolic, or computational.

## GEOMETRIC ENHANCEMENT

In general, there are various kinds of geometric enhancements that may be introduced into spatial displays, but their common feature is a transformation of the metrics of either the displayed space or of the objects it contains. A familiar example is found in relief topographic maps for which it is useful to exaggerate the vertical scale. This technique has also been used for experimental traffic displays for commercial aircraft (fig. 18) (Ellis, McGreevy, and Hitchcock, 1987).

Another type of geometric enhancement important for displays of objects in 3D space involves the choice of the position and orientation of the eye coordinate system used to calculate the projection (fig. 19). Azimuth, elevation, and roll of the system may be selected to project objects of interest with a useful aspect. This selection is particularly important for displays without stereoscopic cues, but all types of displays can benefit from an appropriate selection of these parameters (Ellis et al., 1985; see also Paper 30, Kim et al., 1987).

The introduction of deliberate spatial distortion into a spatial instrument can be a useful way to use geometric enhancement to improve the communication of spatial information to a viewer. The distortion can be used to correct underlying natural biases in spatial judgements. For example, ·exocentric direction judgements (Howard, 1982) made of extended objects in perspective displays, can, for some response measures, exhibit a "telephoto bias." That is to say, the subjects behave as if they were looking at the display through a telephoto lens. This bias can be corrected by introduction of a compensating wide-angle distortion (McGreevy and Ellis, 1986; Grunwald and Ellis, 1987).

## SYMBOLIC ENHANCEMENT

Symbolic enhancements generally consist of objects, scales, or metrics that are introduced into a display to assist pick-up of the communicated information. The usefulness of such symbolic aids can be seen, for example, in displays to present air traffic situation information which focus attention on the relevant "variables" of a traffic encounter, such as an intruder's relative position, as opposed to less useful "properties" of the aircraft state, such as absolute position (Falzon, 1982).

One way to present an aircraft's position relative to a pilot's own ship on a perspective display is to draw a grid at a fixed altitude below an aircraft symbol and drop reference lines from the symbol onto the grid (fig. 20). If all the displayed aircraft are given predictor vectors that show future position, a similar second reference line can be dropped from the ends of the predictor lines.

The second reference line not only serves to clearly show the aircraft the future position of the aircraft on the grid, but additionally clarifies the symbol's otherwise ambiguous aspect. Interestingly, it can also improve perception of the target's heading difference with a pilot's ownship. This effect has been shown in an experiment examining the effects of reference lines on egocentric perception of azimuth (Ellis, Grunwald, and Velger, 1987). I wish to briefly use this experiment

as an example of how psychophysical evaluation of images can help improve their information display effectiveness.

In this experiment subjects viewed static perspective projects of aircraft-like symbols elevated at three different levels above a ground reference grid: a low level below the view vector, a middle level colinear with the viewing vector, and a high level above the view vector. The aircraft symbols had straight predictor vectors projecting forward, showing future position. In one condition, reference lines were dropped only from the current aircraft position; in the second, condition lines were dropped from both current and predicted position.

The first result of the experiment was that subjects made substantial errors in their estimation of the azimuth rotation of the aircraft; they generally saw it rotated more towards their frontal plane than it in fact was. The second result was that the error towards the frontal plane for the symbols with one reference line increased as the height of the symbol increased above the grid. Most significantly, however, introduction of the second reference line totally eliminated the effect of height, reducing the azimuth error in some cases almost 50% (fig. 21).

More detailed discussion of this result is beyond the scope of this talk; however, these experimental results show in a concrete way how appropriately chosen symbolic enhancements can provide not only qualitative, but quantitative, improvement in pictorial communication. They also show that appropriate psychophysical investigations can help designers define their spatial instruments.

## COMBINED GEOMETRIC AND SYMBOLIC ENHANCEMENTS

Some enhancements combine both symbolic and geometric elements. One interesting example is provided by techniques connecting the photometric properties of objects or regions in the display with other geometric properties of the objects or regions themselves. Russell and Miles (1987) (see also Paper 48), for example, have controlled the transparency of points in space with the gradient of the density of a distributed component and produced striking visualization of 3D objects otherwise unavailable. These techniques have been applied to data derived from sequences of MRI or CAT scans and allowed a kind of "electronic dissection" of medical images. Though these techniques can provide absolutely remarkable images, one of the challenges of their use is the introduction of metrical aids to allow the viewer to pick up quantitative information from the photometric transformation (Meagher, 1985, 1987).

## COMPUTATIONAL ENHANCEMENTS

While considerable computation may be involved in the rendering and shading of static pictures, the importance of computational enhancement is also particularly evident for shaping and placing objects in interactive spatial instruments. In principle, if unlimited computational resources were available, no computational enhancements would be needed. The enhancements are necessary because resources must be allocated to ensure that the image is computed in a timely and appropriate manner.

An example of a computational enhancement can be found in the selection of a type of geometric distortion to use as a geometric enhancement in a head-mounted, virtual-image computer display of the type pioneered by Ivan Sutherland (1970) (fig. 22). Distortions in the imagery used by such displays can be quite useful, since they are one way that the prominence of the components of the image could be controlled.

It is essential, however, that the enhancements operate on the displayed objects before the viewing transformation, because, here the picture of reality collides with a picture for communication. The virtual-image presentation makes the picture appear in some ways like a real space. Accordingly, distorting geometric enhancements that are computed after the viewing transformation can disturb visual-vestibular coordination and produce nausea and disorientation. This disturbance shows how different computational constraints distinguish head-mounted from panel-mounted formats.

A second example of a computational enhancement is shown on the interactive, proximity-operations, orbital planning tool developed by Art Grunwald in our laboratory. When first implemented, the user was given control of the direction and magnitude of the thrust vector; these seemed reasonable, since they are the basic inputs to making an orbital change. The nonlinearities and counterintuitive nature of the dynamics, however, made manual control of a predictor cursor driven by these variables impossible. The computational trick needed to make the display tool work was allowing the user to command that the craft be at a certain location at a set time and allow the computer to calculate the required burns through an inverse orbital dynamics algorithm. This technique provided a good match between the human user's planning abilities and the computer's massive computational capacity.

A third example of a computational enhancement is shown on the same interactive, proximity-operations, orbital planning tool. Despite the fact that the system has been implemented on a high-performance 68020 workstation with floating-point processor and dedicated graphics geometry engine, unworkably long delays would occur if the orbital dynamics were constantly updated while the user adjusted the cursor to plan a new way-point. Accordingly, the dynamics calculations are partially inhibited whenever the cursor is in motion. This feature allows a faster update when the user is setting a way-point position and eliminates what would otherwise be an annoying delay of about 0.3 sec while adjusting the way-point position.

When Arthur Grunwald finished the first iteration of this display, we decided to name it. Like a dutiful NASA researcher, he searched for a acronym—something like Integrated Orbital and Proximity Planning Systems, or IOPPS for short. This looked to me like it might sound like OOPS and I thought we should find a better name. I asked him to find maybe a Hebrew name that would be appropriate. He thought about it for awhile and came up with Navie, or "reliable prophet." This is perfect, since that is exactly what the display is intended to provide: reliable prophesy of future position.

But there is another sense in which Navie is a good name. I would like to think that it, and other display concepts developed in our division and elsewhere, also provide a kind of prophesy for the coming displays to be used by NASA during future unmanned, and manned, exploration of air and space.

Like most human activities, this exploration is not an endeavor that can be automated; it will require iteration, trial and error, interactive communication between men and machines and between men and other men. The media for this communication must be designed. Some of them will be spatial instruments.

# BIBLIOGRAPHY AND REFERENCES

Adams, A. (1975). Camera and lens. New York: Morgan and Morgan.

Bertoz, A. and Melville-Jones, G. (1985). Adaptive mechanisms in gaze control: facts and theories. New York: Elsevier.

Boeing (1983). 757/757 Flight deck design development and philosophy, D6T11260-358. Boeing, Seattle.

Bunge, W. (1965). Theoretical Geography, 2nd ed. Studies in Geography, The Netherlands, Lund, Gieerup.

De Solla Price, Derek J. (1959). An ancient greek computer. Sci. Amer., 200, 60-67.

Dickinson, G. C. (1979). Maps and air photographs. New York: Wiley.

Ellis, S. R., Kim, W. S., Tyler, M., McGreevy, M. W., and Stark, L. (1985). Visual enhancements for perspective displays: perspective parameters. Proc. Intern. Conf. Systems, Man, and Cybernetics. IEEE Catalog 85CH2253-3, 815-818.

Ellis, S. R., Kim, W. S., Tyler, M., and Stark, L. (1985). In Proc. 1985 Intern. Conf. Systems, Man, and Cybernetics. New York: IEEE. 815-818.

Ellis, S. R., Grunwald, A., and Velger, M. (1987). Head-mounted spatial instruments: synthetic reality or impossible dream. Proc. 1987 AGARD Symp. "Motion cues in flight simulation and simulator induced sickness," Brussels, Belgium.

Ellis, S. R., McGreevy, M. W., and Hitchcock, R. (1987). Perspective traffic display format and airline pilot traffic avoidance. Human Factors, 29, 371-382.

Ellis, S. R., Smith, S., and McGreevy, M. W. (1987). Distortions of perceived visual directions out of pictures. Perception Psychophys., 42, 535-544.

Ellis, S. R. and Grunwald, A. (1987). A new visual illusion of projected three-dimensional space. NASA TM 100006.

Eyles, D. (1986). Space stations thrillers unfold at Draper Lab. Aerosp. amer., 24, 38-41.

Falzon, P. (1982). Display structures: compatibility with the operators mental representations and reasoning processes. Proc. 2nd European Ann. Conf. Human Decision Making and Manual Contr. Wachtberg-Werthoven, Federal Republic of Germany: Forschungsinstitut für Anthropotechnik. Pp. 297-305.

Fisher, S. S., McGreevy, M. W., Humphries, J., and Robinett, W. (1986). Virtual environment display system. ACM 1986 Workshop on 3D Interactive Graphics, Chapel Hill, NC.

Foley, J. D. and Van Dam, A. (1982). Fundamentals of interactive computer graphics. Boston: Addison-Wesley.

Gombrich, E. H. (1969). Art and illusion, Princeton, N. J.: Princeton Univ. Press.

Gregory, R. L. (1970). The intelligent eye. New York: McGraw-Hill.

Held, R., Efstathiou, A., and Greene, M. (1966). Adaptation to displaced and delayed visual feedback from the hand. J. Exp. Psychol., 72, 887-891.

Ittelson, W. H. (1951). Size as a cue to distance: static localization. Amer. J. Psychol., 64, 54-67.

Goldstein, E. B. (1987). Spatial layout, orientation relative to the observer, and perceived projection in pictures viewed at an angle. J. Exp. Psychol., 13, 256-266.

Grunwald, A. and Ellis, Stephen R. (1986). Spatial orientation by familiarity cues. Proc. 6th European Ann. Conf. Manual Contr., Univ. Keele, Great Britain.

Grunwald, A. and Ellis, S. R. (1987). Interactive orbital proximity operations planning system. NASA TP-2839.

Helmholtz, H. Handbook of physiological optics (1956-1866). Southall trans., Opt. Soc. Amer. (1924), Rochester, NY.

Herbst, P. J., Wolff, D. E., Ewing, D., and Jones, L. R. (1946). The TELERAN proposal. Electronics, 19, 125-127.

Howard, I. (1982). Human visual orientation. New York: Wiley.

Jenks, G. F. and Brown, D. A. (1966). Three-dimensional map construction. Science, 154, 857-846.

Kim, W. S., Ellis, S. R., Tyler, M., and Hannaford, B. (1987). IEEE Trans. Man and Cybernetics SMC-17, no. 1, pp. 61-71.

King, H. C. (1978). Geared to the Stars. Toronto: Univ. Toronto Press.

McGreevy, M. W. and Ellis, S. R. (1986). The effects of perspective geometry on judged direction in spatial information instruments. Human Factors, 28, 421-438.

Meagher, D. J. (1985). Surgery by computer. New Scientist, p. 21.

Meagher, D. J. (1987). Manipulation analysis and display of 3D medical objects using octtree encoding. Innov. Tech. Biol. Med., 8, n° special 1.

Muybridge, Edward (1975). Animals in Motion, Lewis S. Brown, ed. New York: Dover.

Nagata, S. (1986). How to reinforce perception of depth in single two-dimensional pictures. Selected papers in basic researchers, No. 44-51, Proc. 1984 SID, 25, 239-246.

Parker, D. E., Renschke, M. F., Arrott, A. P., Homick, J., and Lichtenberg, B. (1986). Otolyth tilt-translation reinterpretation following prolonger weightlessness: implications for preflight training. Avia., Space, Environ. Med., 56, 601-606.

Piaget, J. and Inhelder, B. (1956). The child's conception of space. London: Routledge and Kegan Paul.

Roscoe, S. N. (1984). Judgements of size and distance with imaging displays. Human Factors, 26, 617-629.

Roscoe, S. N. (1987). The trouble with HUDs and HMDs. Human Factors Soc. Bull., 30, 7, 1-3.

Russell, G. and Miles, R. B. (1987). Display and perception of 3-D space-filling data. Appl. Opt., 26, 973-982.

Sutherland, I. E. (1970). Computer displays. Sci. Amer., 222, 57-81.

Weintraub, D. J., Haines, R. F., and Randle, R. J. (1985). Head up display: HUD utility II: runway to HUD. Proc. 29th Meeting Human Factors Soc. Univ. Michigan, Ann Arbor.

Figure 1.– Prehistoric cave painting of animals from southwestern France.



Figure 2.– Woodcut by Dürer illustrating how to plot lines of sight with string in order to make a correct perspective projective.
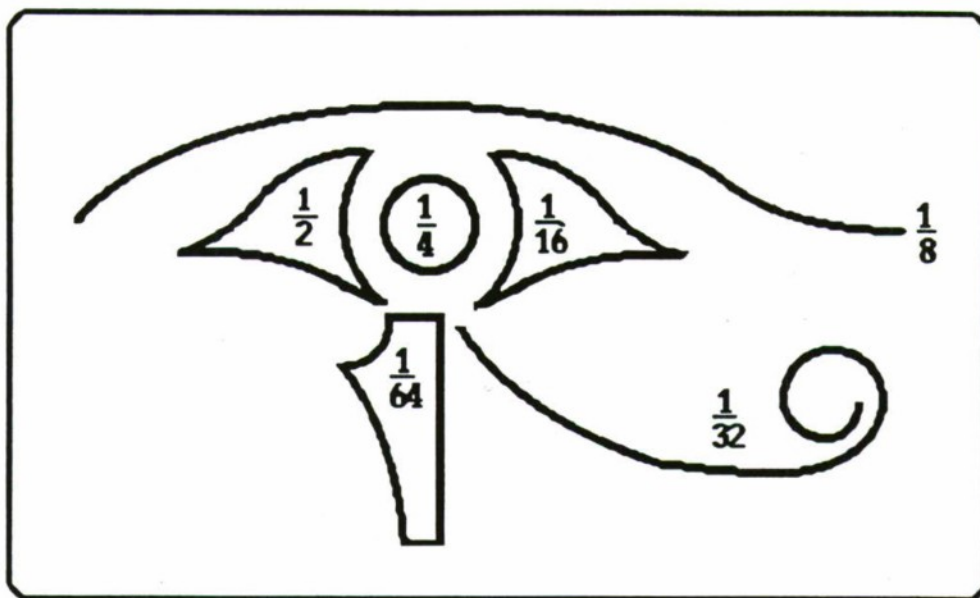
Figure 3.– Egyptian hieroglyphic for the Eye of Horus illustrating the symbolic aspect of pictographs. Each part of the eye is also a symbol for a commonly used fraction. These assignments follow from a myth in which the Sun, represented by the eye, was torn to pieces by the God of Darkness later to be reassembled by Thoth, the God of Learning.



Figure 4.– Leonardo's sketch of two hands using shading to depict depth.

Figure 5.– Crivelli's Annunciation illustrating strong perspective convergence associated with wide-angle views that can exaggerate the range of depth perceived in a picture.
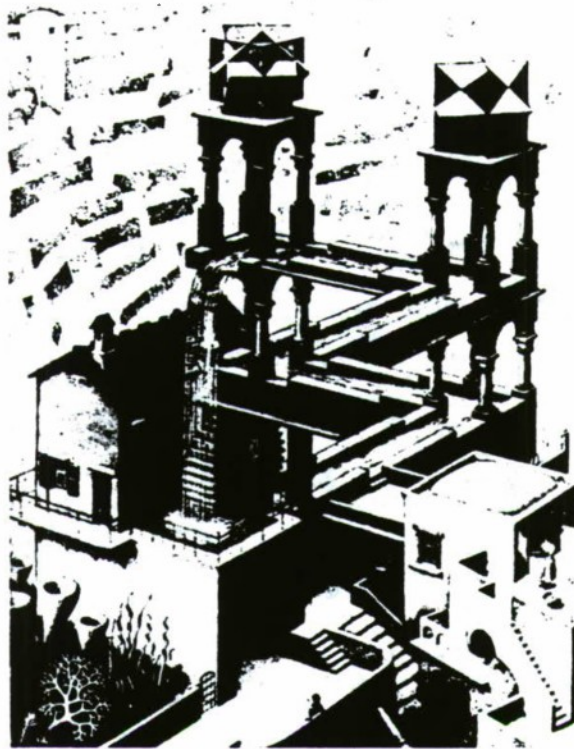
Figure 6.– An engraving by Escher illustrating how the ambiguity of depicted height and depicted depth can be used in a picture to create an impossible structure, apparently allowing water to run uphill. © 1988 M. C. Escher heirs/Cordon Art-Baarn-Holland.

Figure 7.– Urban freeways, a painting by Thiebaud showing an instant of time on a California freeway.
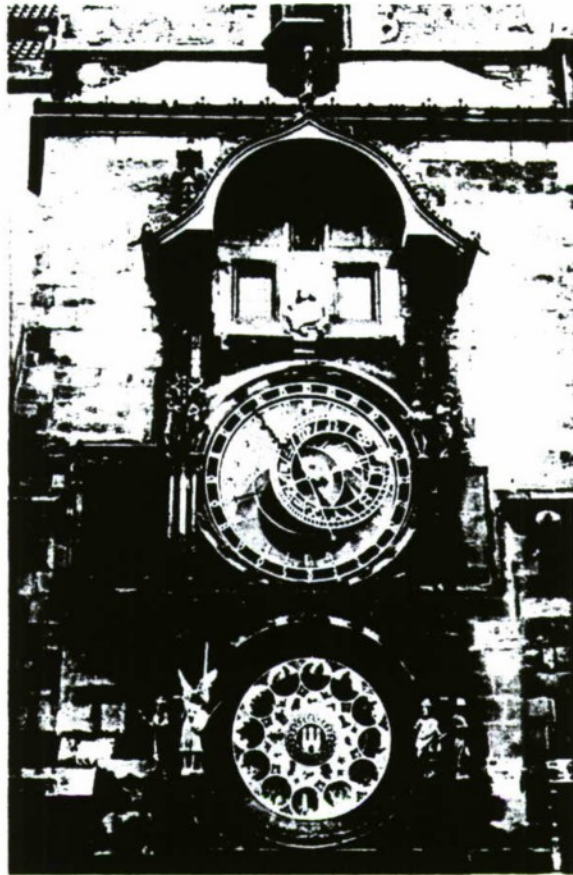
Figure 8.– View of the Prague town hall clock, which indicates the positions of heavenly bodies as well as the time.
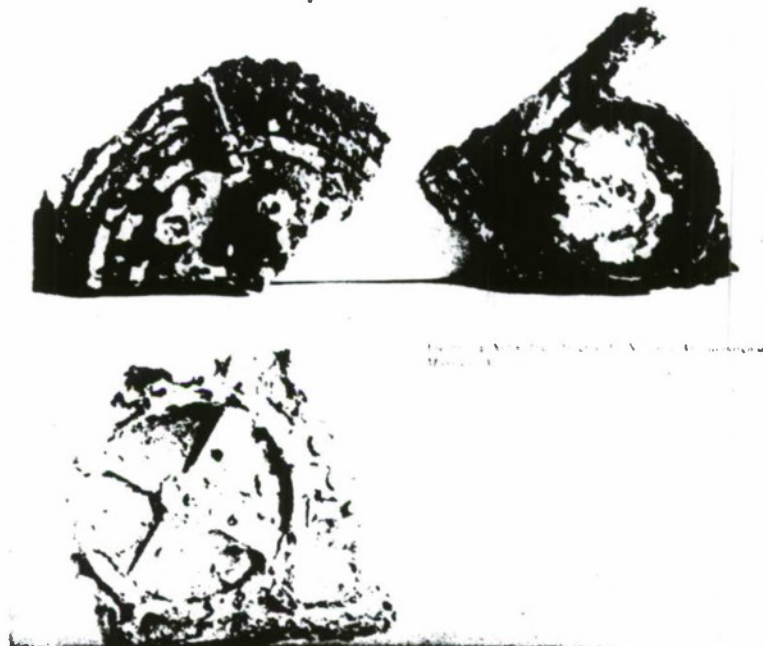
Figure 9.– Fragments of an ancient Greek mechanical device used to calculate the display positions of heavenly bodies.
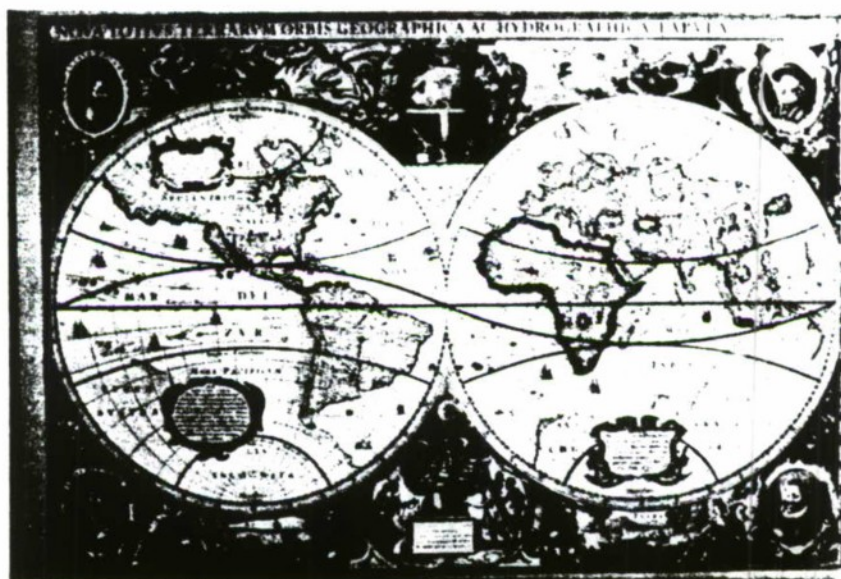


Figure 10.– An old map of the world from the 17th Century.

Figure 11.– Rhumb-line and great-circle routes between two points on the globe. Note the constant bearing of the rhumb-line route and the constantly changing bearing of the great-circle route. On the globe the great-circle route is analogous to a straight line and direction Z is the azimuth of B from A.



Figure 12.– Plate caree projection illustrating the curved path traced by a rhumb line on this format, i.e., line AEFG.
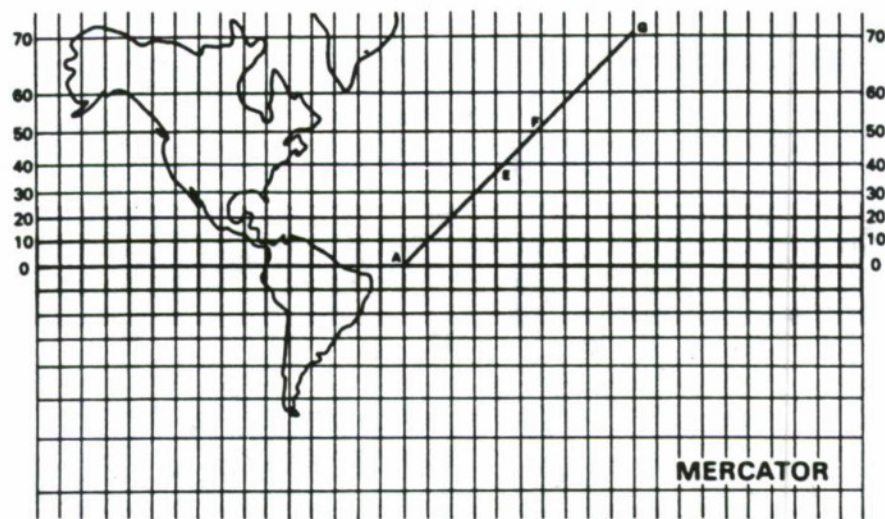
Figure 13.– Mercator projection illustrating how a nonlinear distortion of the latitude scale can be used to straighten out the path traced by a rhumb line.
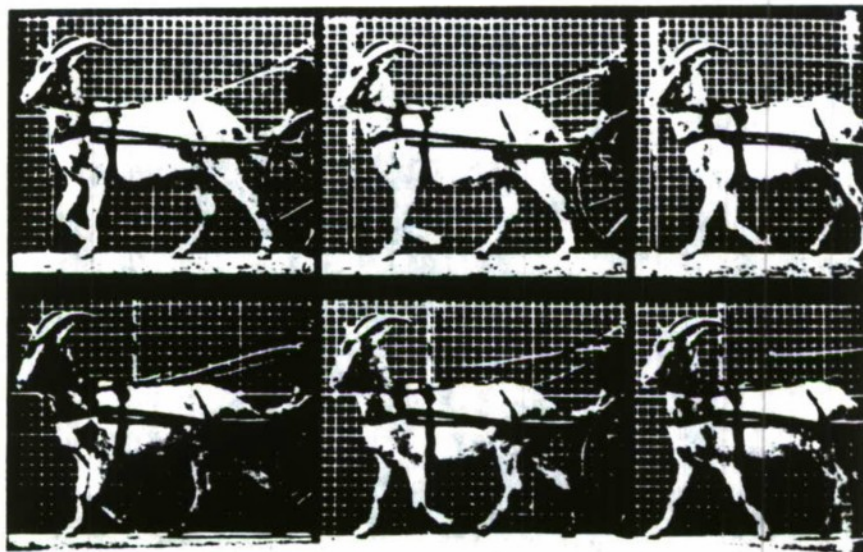


Figure 14.– Muybridge's photographic sequence of a goat walking. The background grid provides a reference for measuring the pattern of limb movement.
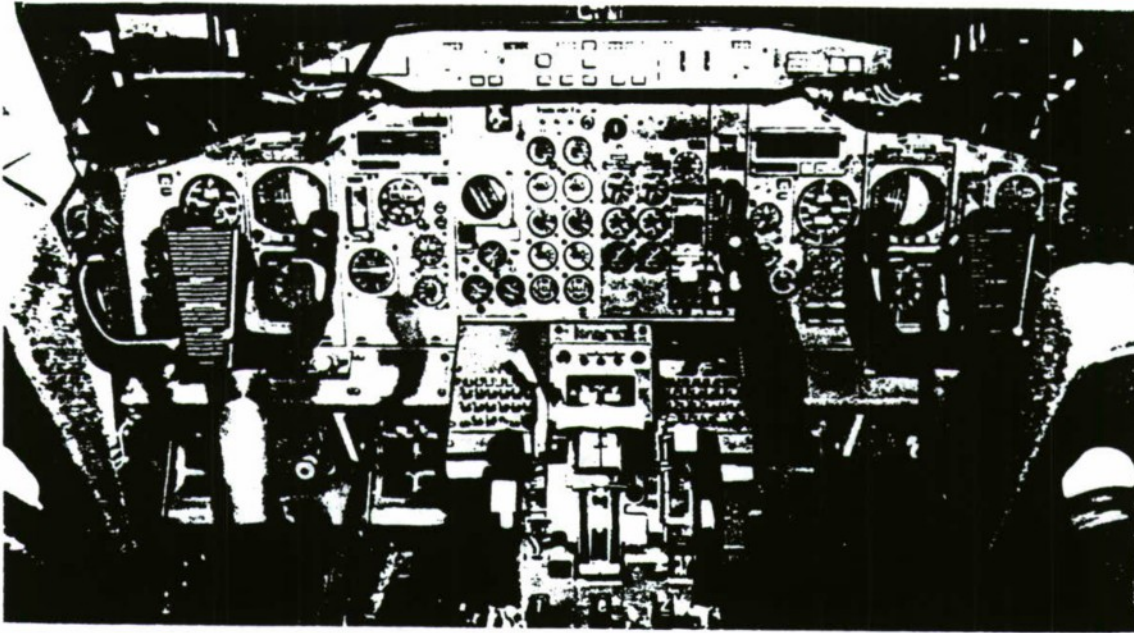
Figure 15.– View of the forward panel of a 737 cockpit showing the artificial horizon on the attitude direction indicator.
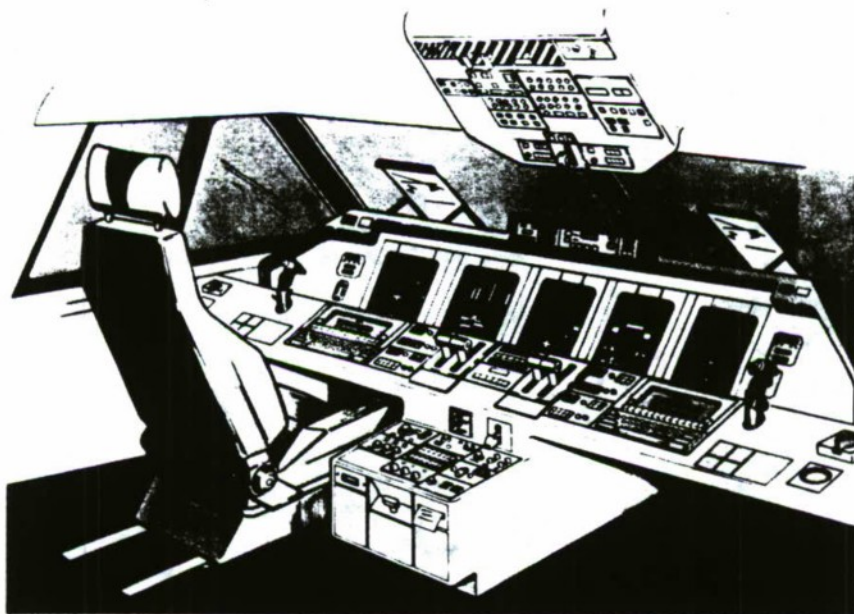


Figure 16.– An advanced-concepts commercial aircraft cockpit in the Man-Vehicle Systems Research Facility of NASA Ames Research Center. This artist's conception shows how future cockpits may resemble ordinary offices.
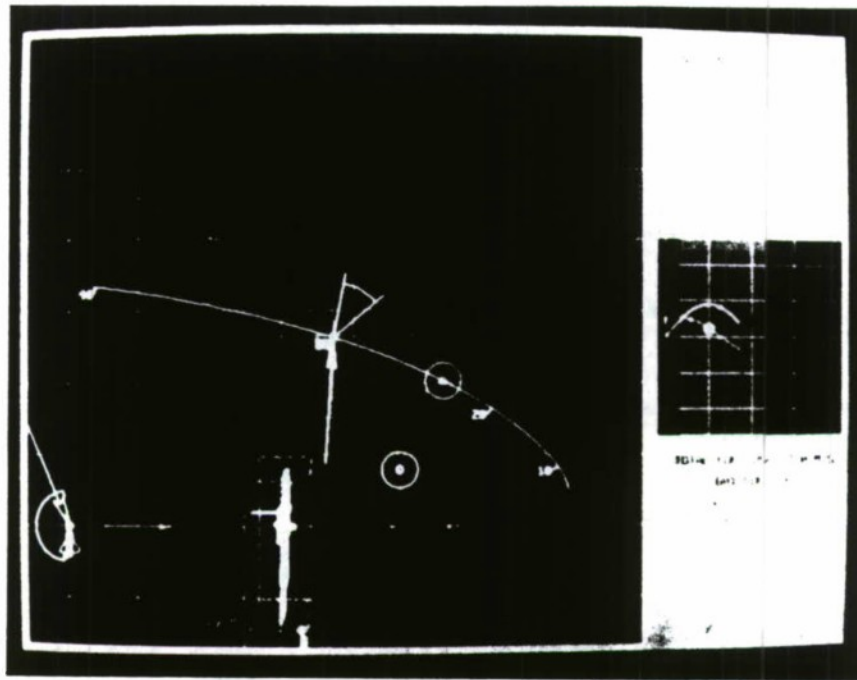
Figure 17.– Sample view from an interactive-graphics-based, planning tool to be used in assisting informal changes in orbits and proximity operations in the vicinity of a space station.
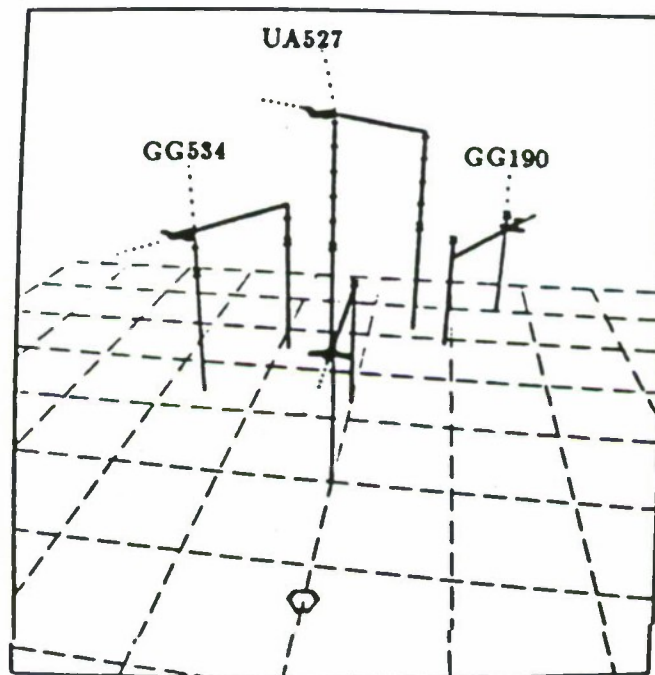


Figure 18.– Possible display format for a commercial aircraft cockpit traffic display. The pilot's own craft is shown in the center of the display. All aircraft have predictor vectors attached showing future position and have reference lines to indicate height above a reference grid.
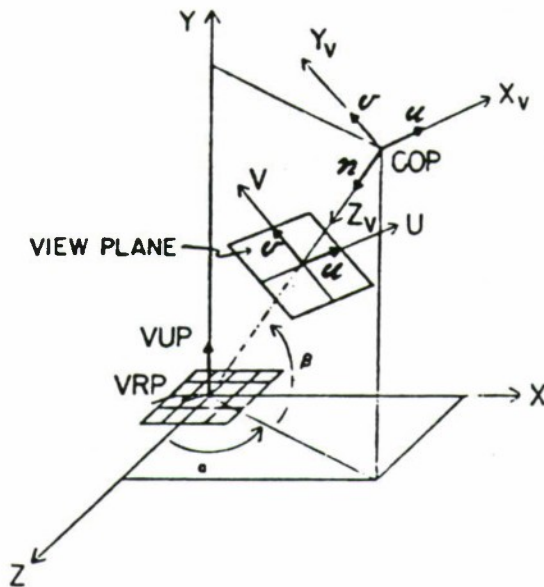
Figure 19.– Illustration of the geometry of perspective projection showing the azimuth and the elevation of the viewing vector InR, directed from the center of projection COP.
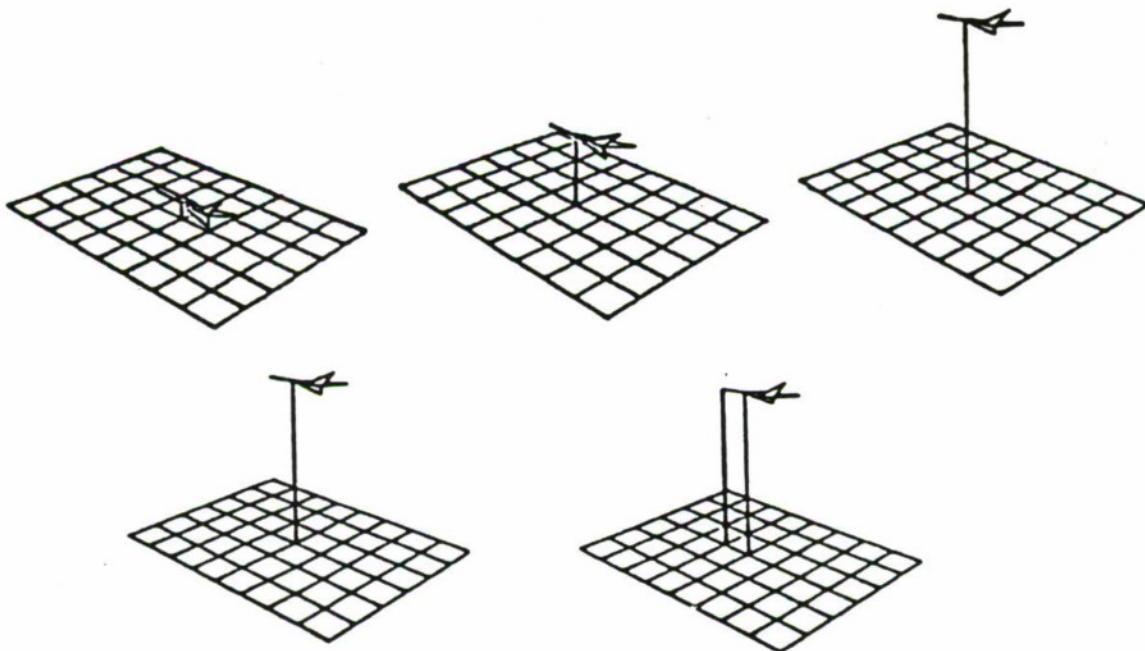


Figure 20.– Five views of sample stimuli used to examine the perceptual effect of raising an aircraft symbol above a reference grid. The attitude of the symbol is kept constant. Addition of a second vertical reference line is seen to reduce the illusory rotation caused by the increasing height above of the grid.
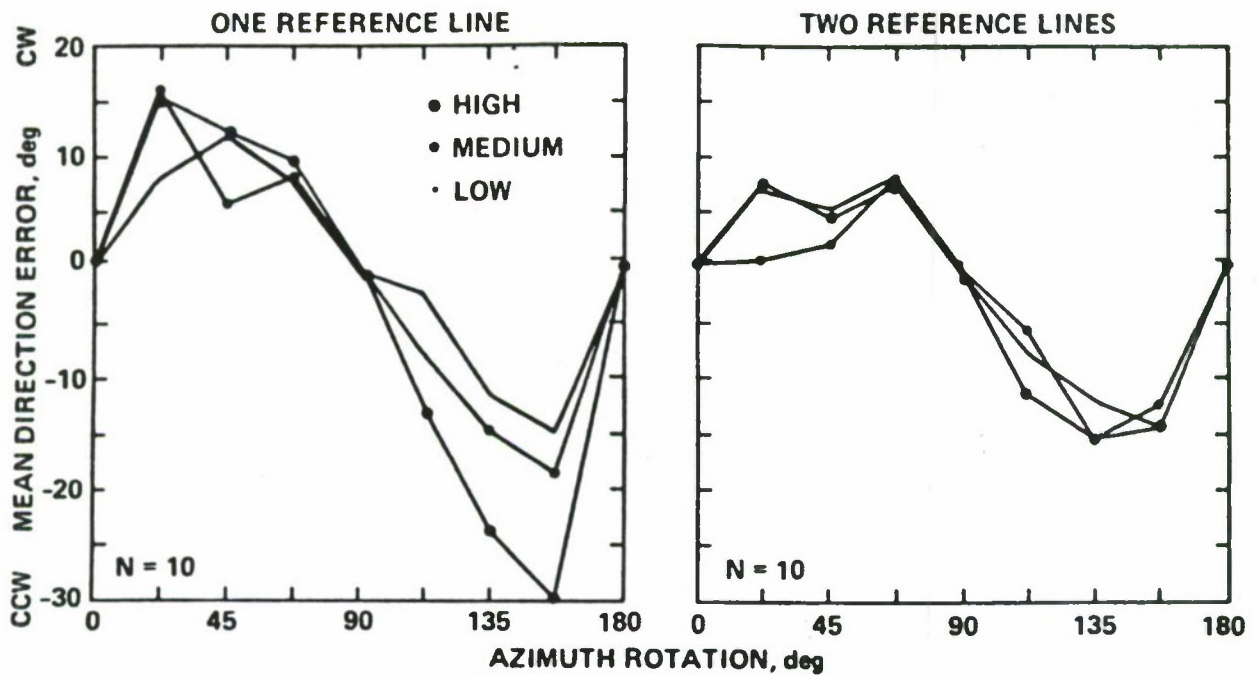
Figure 21.– Mean clockwise and counterclockwise egocentric direction judgement for clockwise azimuth rotation of an aircraft symbol.
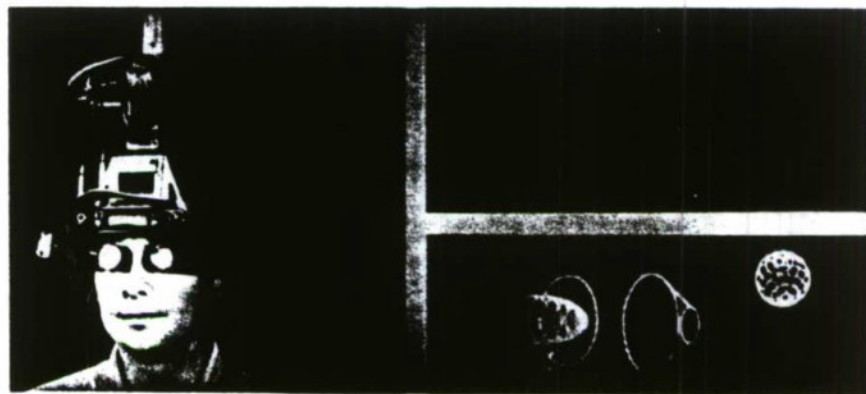


Figure 22.– Probably the first computer-driven head-mounted viewing device. It was developed by Ivan Sutherland to give the viewer the illusion of actually being in the synthetic world defined in the computer.

# SPATIAL PERCEPTION: PRIMARY DEPTH CUES

# SPATIAL CONSTRAINTS OF STEREOPSIS IN VIDEO DISPLAYS

Clifton Schor
University of California
School of Optometry
Berkeley, California

Recent development in video technology, such as the liquid crystal displays and shutters, have made it feasible to incorporate stereoscopic depth into the three-dimensional representations on two-dimensional displays. However, depth has already been vividly portrayed in video displays without stereopsis using the classical artists' depth cues described by Helmholtz (1866) and the dynamic depth cues described in detail by Ittleson (1952). Successful static depth cues include overlap, size, linear perspective, texture gradients, and shading. Effective dynamic cues include looming (Regan and Beverly, 1979) and motion parallax (Rogers and Graham, 1982).

Stereoscopic depth is superior to the monocular distance cues under certain circumstances. It is most useful at portraying depth intervals as small as 5-10 arc seconds. For this reason it is extremely useful in user-video interactions such as in telepresence. Objects can be manipulated in 3-D space, for example, while a person who controls the operations views a virtual image of the manipulated object on a remote 2-D video display. Stereopsis also provides structure and form information in camouflaged surfaces such as tree foliage. Motion parallax also reveals form; however, without other monocular cues such as overlap, motion parallax can yield an ambiguous perception. For example, a turning sphere, portrayed as solid by parallax, can appear to rotate either leftward or rightward. However, only one direction of rotation is perceived when stereo-depth is included. If the scene is static, then stereopsis is the principal cue for revealing the camouflaged surface structure. Finally, dynamic stereopsis provides information about the direction of motion in depth (Regan and Beverly, 1979). When optical flow patterns seen by the two eyes move in phase, field motion is perceived in the fronto-parallel plane. When optical flow is in antiphase (180°) motion is seen in the saggital plane. Binocular phase disparity of optical flow as small as 1° can be discriminated as changes in visual direction of motion in a 3-D space (Beverly and Regan, 1975). This would be a useful addition to the visual stimuli in flight simulators.

Several spatial constraints need to be considered for the optimal stimulation of stereoscopic depth. The stimulus for stereopsis is illustrated in figure 1. Each peg subtends a visual angle at the entrance pupils of the eyes, and this angle is referred to as binocular parallax. The difference in this angle and the angle of convergence forms an absolute disparity. In the absence of monocular depth cues, perceived distance of an isolated target, subtending an absolute disparity is biased toward 1.5 meters from the physical target distance. Gogle and Teitz (1973) referred to this as equidistance tendency. If the target moves abruptly from one distance to another, convergence responses signal the change of depth (Foley and Richards, 1972); however, smooth continuous changes in binocular parallax, tracked by vergence eye movements do not cause changes in perceived distance (Erkelens and Collewijn, 1985; Guttmann and Spatz, 1985). Once more than one disparate feature is presented in the field, differences in depth (stereopsis), stimulated by retinal image disparity become readily apparent. Stereothresholds may be as low as 2 sec arc, which ranks stereopsis along with vernier and bisection tasks among the hyperacuities.

Stereo-sensitivity to a given angular depth interval varies with the saggital distance of the stimulus depth increment from the fixation plane. Sensitivity to depth increments is highest at the horopter or fixation plane where the disparity of one of the comparison stimuli is zero (Blakemore, 1970). This optimal condition for stereopsis was used by Tschermack (1930) as one of four criteria for defining the empirical longitudinal horopter. The Weber fraction describing the ratio of increment stereothreshold (arc sec) over the disparity pedestal (arc min) (3 sec/min) is fairly constant with disparity pedestal amplitudes up to 1°. This fraction was derived from figure 2, which plots stereothreshold in seconds of arc at different saggital distances in minutes arc from the fixation point for targets consisting of vertical bars composed of coarse or fine features. A two-alternative, forced choice is used to measure a just-noticeable difference between a depth increment between an upper test bar and a lower standard bar, both seen at some distance before or behind the fixation plane. The bar used was a narrow-band, spatially filtered line produced from a difference of Gaussians (DOG) whose center spatial frequency ranges from 9.5 to 0.15 cycles/deg (Badcock and Schor, 1985). When these thresholds are plotted, the slopes of these functions found with different width DOGs are the same on a logarithmic scale. However, thresholds for low spatial frequencies (below 2.5 cpd) are elevated by a constant disparity which illustrates they are a fixed multiple of thresholds found with higher spatial frequencies. These results illustrate that depth stimuli should be presented very near the plane of fixation, which is the video screen.

Stereo-sensitivity remains high within the fixation plane over several degrees about the point of fixation. Unlike the rapid reduction of stereo-sensitivity with overall depth or saggital distance from the horopter, stereo-sensitivity is fairly uniform and at its peak along the central 3° of the fixation plane (Blakemore, 1970; Schor and Badcock, 1985). Figures 2 and 3 illustrate a comparison of stereo-depth increment sensitivity for this fronto-parallel stereo and the saggital off-horopter stereothreshold. Also plotted in figure 3 are the monocular thresholds for detecting vernier offset of the same DOG patterns at the same retinal eccentricities. Clearly, stereopsis remains at its peak at eccentricities along the horopter and there is a percipitous fall of visual acuity (Wertheim, 1894) and, as shown here, of vernier acuity over the same range of retinal eccentricities where stereo increment sensitivity is unaffected (Schor and Badcock, 1985). Thus, stereoacuity is not limited by the same factors that limit monocular vernier acuity because the two thresholds differ by a factor of 8 at the same eccentric retinal locus.

In addition to the threshold or lower disparity limit (LDL) for stereopsis, there is an upper disparity limit (UDL), beyond which stereo depth can no longer be appreciated. This upper limit is small, being approximately 10 arc min with fine (high-frequency) targets, and somewhat larger (several degrees) with coarser (low spatial frequency) fusion stimuli (Schor and Wood, 1983). This depth range can be extended either by briefly flashing targets (Westheimer and Tanzman, 1956) or by making vergence movements between them (Foley and Richards, 1972) to a UDL of approximately 24°. The UDL presents a common pitfall for many stereo-camera displays that attempt to exaggerate stereopsis by placing the stereo-cameras far apart. Paradoxically, this can produce disparities that exceed the UDL and results in the collapse of depth into the fronto-parallel plane.

Diplopia is another problem that accompanies large disparities. The diplopia threshold is slightly smaller than the UDL for static stereopsis, and depth stimulated by large flashed disparities is always seen diplopically. Normally, this diplopia can be minimized by shifting convergence from one target to another. However, this is not as easily done with a stereo-video monitor. In real space the stimulus for vergence is correlated with the stimulus for accommodation. With video displays, the stimulus for accommodation is fixed at the screen plane while vergence is an

independent variable. Because there is cross-coupling between accommodation and vergence, we are not completely free to dissociate these motor responses (Schor and Kotulak, 1986). With some muscular effort, a limited degree of vergence can be expected while accommodation is fixed, depending on the accommodative-convergence ratio (AC/A). When this ratio is high, a person must choose between clearness and singleness.

Additional problems for stereoscopic depth occur with abstract scenes containing high spatial frequency surface texture. This presents an ambiguous stimulus for stereopsis and fusion which can have an enormous number of possible solutions as illustrated by the wallpaper illusion or by a random-dot stereogram. The visual system uses various strategies to reduce the number of potential fusion combinations and certain spatial considerations of targets presented on the visual display can help implement these strategies. A common technique used in computer vision is the coarse-to-fine strategy. The visual display is presented with a broad range of spatial frequency content. The key idea here is that there is little confusion or ambiguity with coarse features like the frame of a pattern. These can be used to guide the alignment of the eyes into registration with finer features that present small variations in retinal image disparity. Once in registration, small disparities carried by the fine detail can be used to reveal the shape or form of the depth surface. An essential condition for this algorithm to work is that sensitivity to large disparities be greatest when they are presented with coarse detail and that sensitivity to small disparities be highest with fine (high spatial frequency) fusion stimuli. This size-disparity correlation has been verified for both the LDL and UDL by Schor and Wood (1983). Figure 4 illustrates the variation of stereothreshold (LDL) and the UDL with spatial frequency for targets presented on a zero disparity pedestal at the fixation point. Stereothresholds are lowest and remain relatively constant for spatial frequencies above 2.5 cycles/deg. Thresholds increase proportionally with lower spatial frequencies. Even though stereothreshold varies markedly with target coarseness, suprathreshold disparities needed to match the perceived depth of a standard disparity are less dependent on spatial frequency. This depth equivalence constitutes a form of stereo-depth constancy (Schor and Howarth, 1986). Similar variations in the diplopia threshold or binocular fusion limit are found by varying the coarseness of fusion stimuli (Schor, Wood, and Ogawa, 1984b).

Figure 5 illustrates that the classical vertical and horizontal dimensions of Panum's fusion limit (closed and open symbols, respectively) are found with high spatial frequency targets, but the fusion limit increases proportionally with the spatial width of targets at spatial frequencies lower than 2.5 cycles/deg. When measured with high-frequency DOGs, the horizontal radius of PFA (Panum's fusional area) is 15 min; and when measured with low-frequency stimuli, PFA equals a 90° phase disparity of the fusion stimulus.

The increase in Panum's fusion limit appears to be caused by monocular limitations to spatial resolution. For example, if the same two targets that were used to measure the diplopia threshold are both presented to one eye to measure a two-point separation threshold, such as the Rayleigh criterion, then the monocular and binocular thresholds are equal when tested with spatial frequencies lower than 2.5 cpd. At higher spatial frequencies we are better able to detect smaller separations between two points presented monocularly than dichoptically. This difference at high spatial frequencies reveals a unique binocular process for fusion that is independent of spatial resolution. With complex targets composed of multiple spatial frequencies, at moderate disparities such as 20 min arc, a diplopia threshold may be reached with high spatial frequency components while stereopsis and fusion may continue with the low spatial frequency components. An example of this simultaneous perception can be seen with the diplopic pixils in a random dot stereogram whose coarse camouflaged form is seen in vivid stereoscopic depth (Duwaer, 1983).

In addition to target coarseness, there are several other aspects of spatial configuration that influence stereopsis and fusion. The traditional studies of stereopsis, such as those conducted by Wheatstone (1838), mainly consider the disparity stimulus in isolation from other disparities at the same or different regions of the visual field. It is said that disparity is processed locally in this limiting case, independent of other possible stimulus interactions other than the comparison between two absolute disparities to form a relative disparity. However, recent investigations have clearly illustrated that in addition to the local processes, there are global processes in which spatial interaction between multiple relative disparities in the visual field can influence both stereopsis and fusion. Three forms of global interactions have been studied. These are disparity crowding, disparity gradients, and disparity continuity or interpolation. These global interactions appear to influence phenomena such as the variation in size of Panum's fusional area, reductions and enhancement of stereo-sensitivity, constant errors or distortions in depth perception, and resolution of a 3-D form that has been camouflaged with an ambiguous surface texture.

Spatial crowding of visual targets to less than 10 arc min results in a depth averaging of proximal features. This is manifest as an elevation of stereothreshold as well as a depression of the UDL (Schor, Bridgeman, and Tyler, 1983). The second global interaction, disparity gradient, depends upon spacing between disparate targets and the difference in their disparities. (Schor and Tyler, 1981). The disparity gradient represents how abruptly disparity varies across the visual field. The effect of disparity gradients upon the sensory fusion range has been investigated with point targets by Burt and Julesz (1980), and with periodic sinuosidal spatial variations in horizontal and vertical disparity by Schor and Tyler (1981). Both groups demonstrate that the diplopia threshold increases according to a constant disparity gradient as the separation between adjacent fusion stimuli increases. Cyclofusion limits are also reduced by abrupt changes in disparity between neighboring retinal regions (Kertesz and Optican, 1974). Stereothresholds can also be described as a constant disparity gradient. As target separation decreases, so does stereothreshold, up to a limit of 15 arc min separation. Further reduction in separation results in crowding, which elevates the stereothreshold. The UDL is also limited by a constant disparity gradient (fig. 5). As spacing decreases, there is a proportional decrease in the UDL. These gradient effects set two strict limitations on the range of stereopscopic depth that can be rendered by the video display. As crowding increases, the UDL will decrease. The effect is that targets exceeding the UDL will appear diplopic and without depth. For example, a top-down picture of a forest which has trees of uneven height will not be seen as uneven depth if the trees are imaged too closely. To remedy this problem, the depth should be reduced by moving the stereocameras closer together. In the other extreme, a shallow slope will not be seen in depth unless it exceeds the gradient for stereothresholds. Even if it does, it may still not be seen if it extends across the entire visual display. Normally there can be unequal optical errors of the two eyes which produce unequal magnification of the two retinal images. This aniso magnification produces an apparent tilt of the stereoscopic frame reference referred to as the fronto-parallel plane. However, this constant depth error is normally corrected or compensated for perceptually (Morrison, 1977). This perceptual compensation could reduce sensitivity to wide static displays of a shallow depth gradient.

A third form of global interaction is observed under conditions where disparity differences between neighboring regions occur too gradually to be detected, such as in the 3-D version of the Craik-Obrien Cornsweet illusion (fig. 6 by Anstis, Howard, and Rogers, 1978), when stereo patterns are presented too briefly to be processed fully (Ramachandran and Nelson, 1976; Mitchison and McKee, 1985), or when several equally probable, but ambiguous, disparity solutions are presented in a region neighboring an unambiguous disparity solution (Kontsevich, 1986). Under all

of these conditions, the depth percept resulting from the vague disparity is similar to or continuous with the depth stimulated by the more visible portion of the disparity stimulus. This illustrates the principle of depth continuity formulated by Julesz (1971) and restated later by Marr and Poggio (1979), which recently was shown by Ramachandran and Cavanaugh (1985) to include the extension of depth to subjective contours in which no physical contour or disparity exists.

Clearly there are many spatial constraints, including spatial frequency content, retinal eccentricity, exposure duration, target spacing, and disparity gradient, which—when properly adjusted—can greatly enhance stereodepth in video displays.

# REFERENCES

Anstis, S. M., Howard, J. P., and Rogers, B. (1978) A Craik-Cornsweet illusion for visual depths. Vision Res. 18: 213-217.

Badcock, D. R., and Schor, C. M. (1985) Depth-increment detection function for individual spatial channels. J. Opt. Soc. Am. A. 2: 1211-1216.

Beverley, K. I. and Regan, D. (1975) The relation between sensitivity and discrimination in the perception of motion-in-depth. J. Physiol. 249: 387398.

Blakemore, C. (1970) The range and scope of binocular depth discrimination in man. J. Physiol. 211: 599-622.

Burt, P. and Julesz, B. (1980) A disparity gradient limit for binocular fusion. Science 208: 615-617.

Duwaer, A. L. (1983) Patent stereopsis and diplopia in random-dot stereograms. Perception and Psychophysics 33: 443-454.

Erkelens, C. J. and Collewijn, H. (1985) Motion perception during dichoptic viewing of moving random-dot stereograms. Vision Res. 25: 583-588.

Foley, J. M. and Richards, W. (1972) Effects of voluntary eye movement and convergence on the binocular appreciation of depth. Perception and Psychophysics 11: 423-427.

Gogle, W. G. and Teitz, J. D. (1973) Absolute motion parallax and the specific distance tendency. Perception and Psychophysics 13: 284-292.

Guttmann, J. and Spatz, H. (1985) Frequency of fusion and of loss of fusion, and binocular depth perception with alternating stimulus presentation. Perception 14: 5-13.

Helmholtz, H. C. (1866) Handbuch der physiologische optik. Hamburg, Voas.

Ittelson, W. H. (1952) The Ames demonstration in perception. Princeton: Princeton University Press.

Ittelson, W. H. (1960) Visual space perception visual space cues: Binocular stereopsis. 123-127.

Julesz, B. (1971) Foundations of cyclopean perception. Chicago, University of Chicago Press.

Kertesz, A. E. and Optican, L. M. (1974) Interactions between neighboring retinal regions during fusional response. Vision Res. 14: 339-343.

Kontsevich, L. L. (1986) An ambiguous random-dot stereogram which permits continuous change of interpretation. Vision Res. 26: 517-519.

Marr, D. and Poggio, T. (1979) A computational theory of human stereo vision. Proc. Royal Soc. 204: 301-328.

Mitchison, G. J. and McKee, S. P. (1985) Interpolation in stereoscopic matching. Nature 315# 6018: 402-404.

Morrison, L. C. (1977) Stereoscopic localization with the eyes asymmetrically converged. Am J. Optometry & Physiol. Optics 54: 556-566.

Ramachandran, V. S. and Cavanagh, P. (1985) Subjective contours capture stereopsis. Nature 314: 527-530.

Ramachandran, V. S. and Nelson, J. I. (1976) Global grouping overrides point-to-point disparities. Perception 5: 125-128.

Regan, D. and Beverley, K. I. (1979) Binocular and monocular stimuli for motion-in-depth: changing disparity and changing size inputs feed the same motion in depth stage. Vision Res. 19: 1331-1342.

Rogers, B. and Graham, M. (1982) Similarities between motion parallax and stereopsis in human depth perception. Vision Res. 22: 261 270.

Schor, C. M. and Badcock, D. R. (1985) A comparison of stereo and vernier acuity within spatial channels as a function of distance from fixation. Vision Res. 25: 1113-1119.

Schor, C. M., Bridgeman, B., and Tyler, C. W. (1983) Spatial characteristics of static and dynamic stereoacuity in strabismus. Investigative Ophthalmol. Vis. Sci. 24: 1572-1579.

Schor, C. M. and Howarth, P. A. (1986) Suprathreshold stereo-depth matches is a function of contrast and spatial frequency perception.

Schor, C. M, and Kotulak, J. (1986) Dynamic interactions between accommodation and convergence are velocity sensitive. Vision Res. 26: 927-942.

Schor, C. M. and Tyler, C. W. (1981) Spatio-temporal properties of Panum's fusional area. Vision Res. 21: 683-692.

Schor, C. M, and Wood, I. (1983a) Disparity ranqe for local stereopsis as a function of luminance spatial frequency. Vision Res. 23: 1649-1654.

Schor, C. M., Wood, I. C. and Ogawa, J. (1984a) Spatial tuning of static and dynamic local stereopsis. Vision Res. 24: 573-578.

Schor, C. M., Wood, I. and Ogawa, J. (1984b) Binocular sensory fusion is limited by spatial resolution. Vision Res. 24: 661-665.

Tschermak, A. (1930) Beitrage zur physiologishen Optik 111; Raimsinn. In Bethe, A., Bergmann, G.V., Embolen, G. and Ellinger, A: Handbuck der normalen und pathologischen physiologic. Berlin Julius Springer vol. 12, p. 2, pp. 833-1000.

Wertheim, T. (1894) Uber die indirekte sekscharfe. Z. Psychol. Physiol. Sinnesory 7: 172-187.

Westheimer, G. and Tanzman, I. J. (1956) Qualitative depth localized with diplopic images. J. Opt. Soc. Am. 46: 116-117.

Wheatstone, C. (1838) Some remarkable phenomena of binocular vision. Phil. Trans. R. Soc. 128: 371-394.
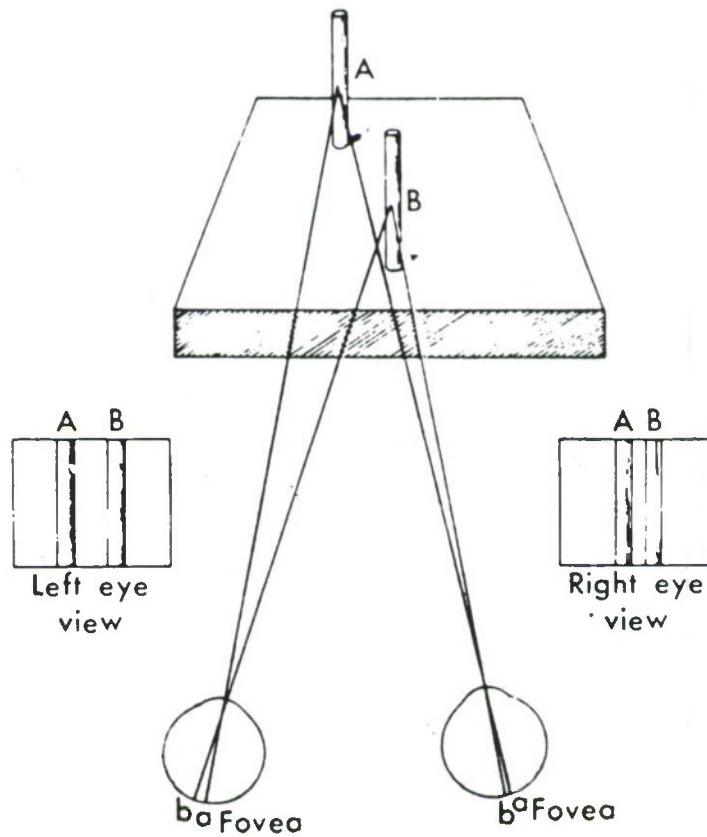
Figure 1. Retinal image disparity based on horizontal separation of the two eyes.
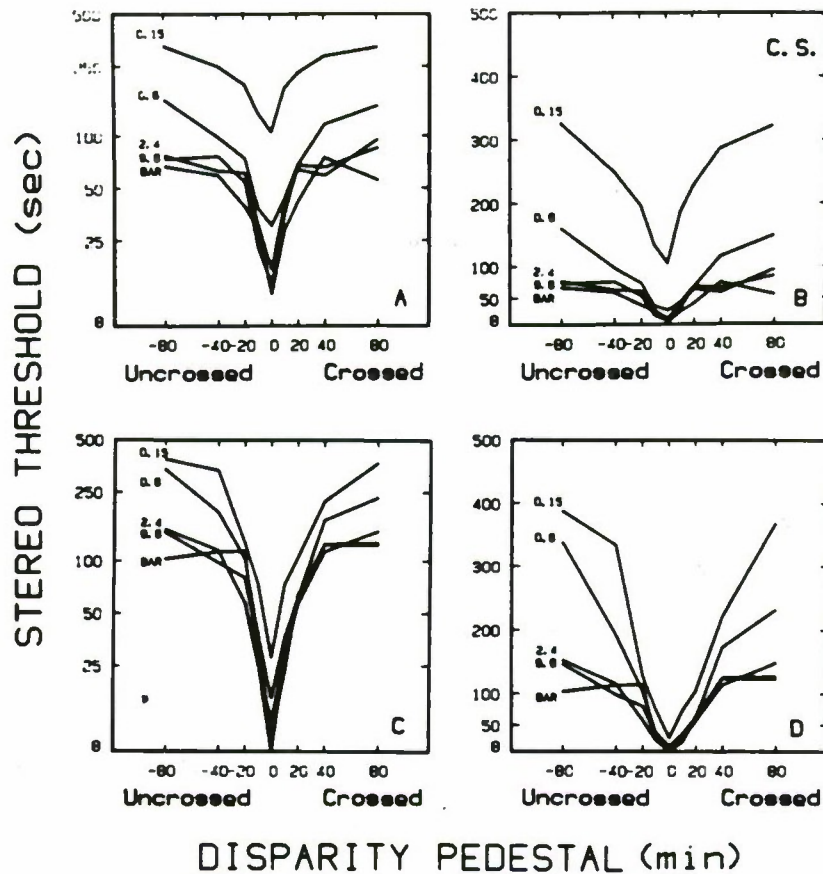
Figure 2. Threshold depth increments obtained, for observer D.B., as a function of pedestal size in both the convergent and divergent directions. Functions illustrate results obtained with a thin bar and DOGs whose center spatial frequencies ranged from 0.15 to 9.6 c/deg. Panels C and D plot the performance measured when the comparison stimulus was a thin bright bar and the test stimulus was a DOG. Panels A and B show the results obtained when a DOG was used both as a comparison and as a test stimulus. Panels A and C plot stereothreshold on a log scale. The data are replotted on a linear scale in panels B and D.
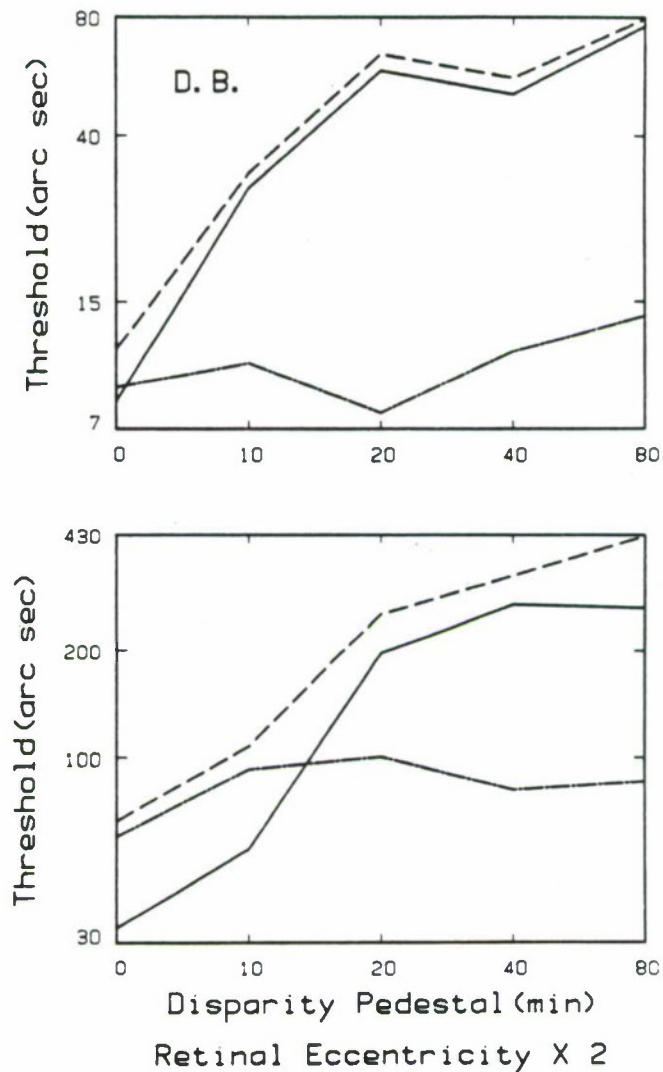
Figure 3. A comparison is made of extra-foveal vernier threshold (solid line) with extra-foveal (mixed dashed line) and extra-horopteral (long dashed line) stereothresholds for a high spatial frequency stimulus (upper plot) and a low spatial frequency stimulus (lower plot). Note that retinal eccentricity has been doubled to be comparable to disparity pedestal. Over a 40 arc min range of retinal eccentricity, stereoacuity remained unchanged and vernier acuity increased moderately. A marked increase in stereothreshold occurred over a comparable (80 arc min) disparity pedestal range.
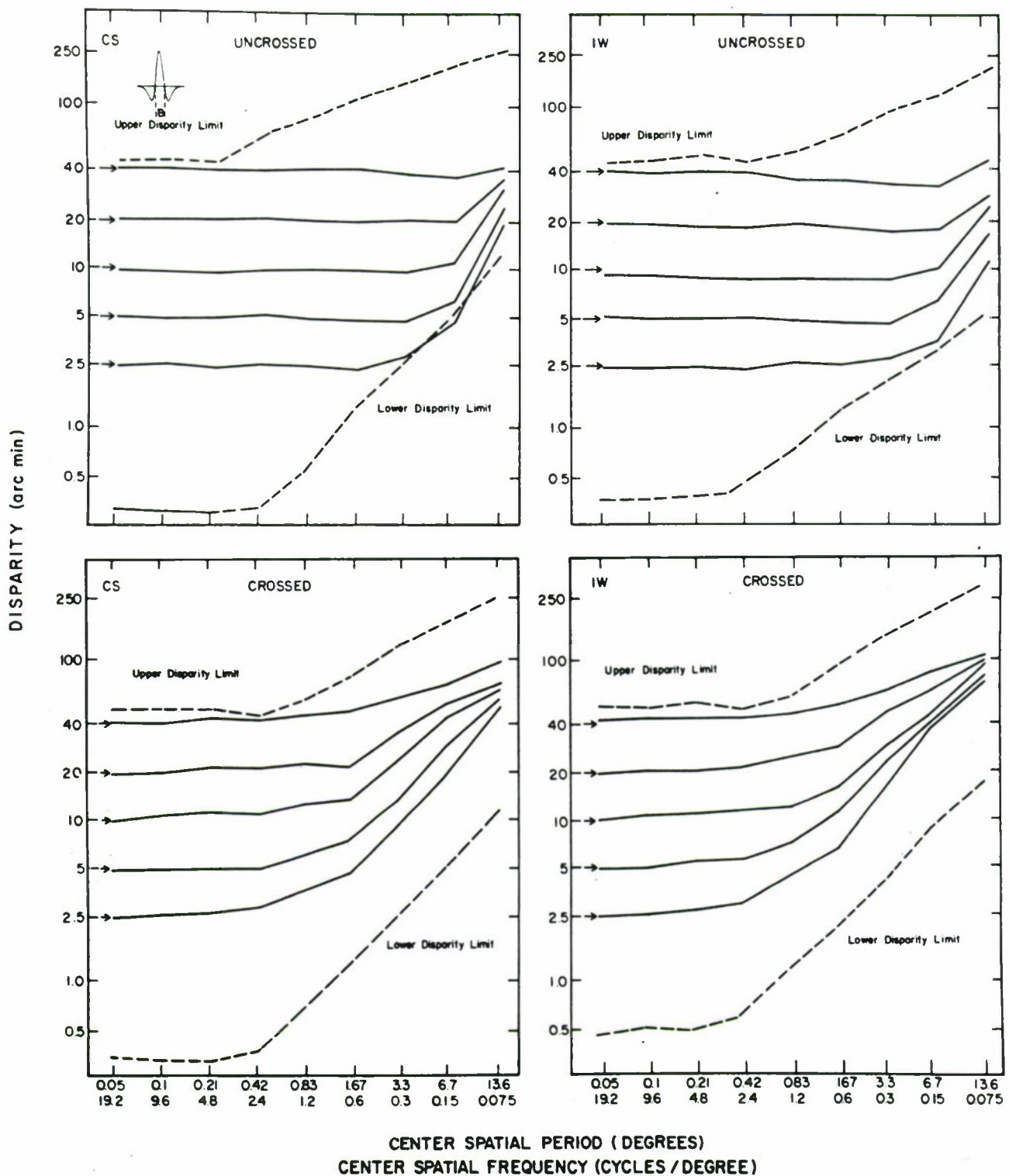
Figure 4. Upper and lower limits for stereopsis are plotted for two subjects as a function of DOG center spatial period along dashed curves at the top and bottom of data sets for uncrossed and crossed disparities respectively. Stereothreshold was lowest at small spatial periods (<0.42 arc min) and increased according to a 6° phase disparity between stereo-half images as spatial period increased. The upper limit increased proportionally to the square root of spatial period over the same range of broad spatial periods. Depth matching curves (solid lines) for several standard suprathreshold disparities (horizontal arrows) have flatter frequency responses than the upper and lower dashed threshold curves. Their breakaway point occurs at a higher spatial period for crossed than for uncrossed disparities. The luminance profile of the difference of two Gaussian functions is inset in the upper left corner.
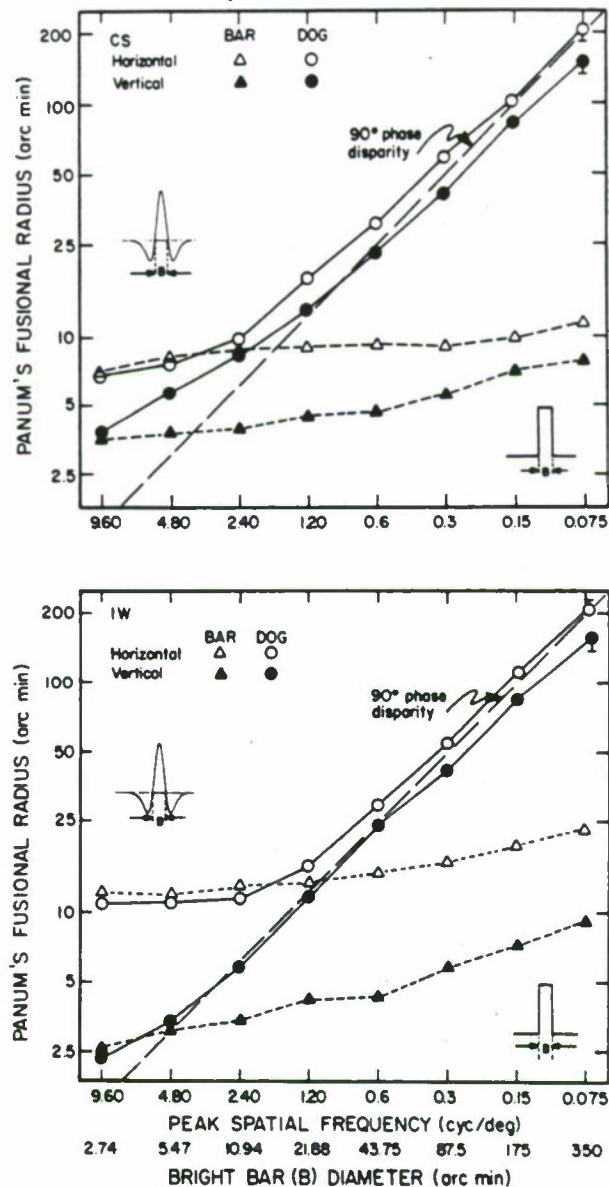
Figure 5. Diplopia thresholds for two subjects are plotted as a function of bright bar width (B) of bar and difference of two Gaussian functions (DOG). Luminance profiles of these two test stimuli are inset below and above the data respectively. A constant phase disparity of 90° is shown by the dashed diagonal line. Horizontal and vertical Panum's fusion ranges (solid lines) coincide with the 90° phase disparity for DOG widths greater than 21 arc min. At the broadest DOG width, the upper fusion limit equals the upper disparity limit for stereoscopic depth perception (bold dashed line). The standard deviation of the mean is shown for the broadest DOG stimulus. At narrow DOG widths, both horizontal and vertical fusion limits approach a constant minimum threshold. Panum's fusion ranges remain fairly constant when measured with bar patterns (dotted lines) and resemble values obtained with high spatial frequency DOGs.
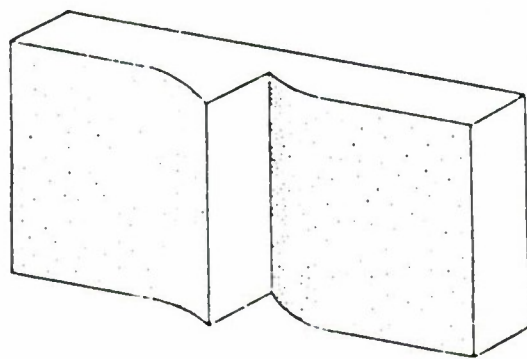
Figure 6. Perspective sketch of the illusory depth surface. Left part looks apparently nearer than the right part.

# STEREOSCOPIC DISTANCE PERCEPTION

John M. Foley
Department of Psychology
University of California
Santa Barbara, California

## INTRODUCTION

Most of this article is concerned with limited cue, open-loop tasks in which a human observer indicates distances or relations among distances. By open-loop tasks I mean tasks in which the observer gets no feedback as to the accuracy of responses. At the end of the article, I will consider what happens when cues are added and when the loop is closed, and what the implications of this research are for the effectiveness of visual displays.

Errors in visual distance tasks do not necessarily mean that the percept is in error. The error could arise in transformations that intervene between the percept and the response. I will argue, however, that the percept is in error. I will argue further that there exist post-perceptual transformations that may contribute to the error or be modified by feedback to correct for the error.

## METHODS

First, I will describe some experiments on binocular distance perception. The stimuli were points of light viewed in dark surroundings. These were in or near the horizontal eye-level plane. The variables that I use are illustrated and defined in figure 1. The angle subtended by straight lines from a stimulus point to the rotation centers of the eyes is the binocular parallax of that point. (It is sometimes called the convergence angle or stimulus to convergence.) The binocular parallax and the horizontal direction, $\theta_i$, serve as coordinates that specify the positions of points in the plane. The binocular disparity of one point relative to another is defined as the binocular parallax of the first, minus the binocular parallax of the second. Note that binocular disparity is a signed quantity; a farther point has a negative disparity relative to a nearer one. The two open dots correspond to the perceived positions of $r$ and $i$. The binocular parallax of the perceived position of a point is called the effective binocular parallax of the point. The difference between two effective binocular parallaxes is an effective binocular disparity. These perceptual variables are defined in the same way as the corresponding physical variables except that perceived distance, $D'$, is substituted for physical distance, $D$, in each equation. I assume that perceived horizontal direction equals physical horizontal direction. There is evidence that this is correct under the conditions of my experiments.

Some of the experiments I will describe were done with stimulus points at different distances. Others were done by simulating the distance dimension stereoscopically. If the stimulus to vergence is not grossly different than the stimulus to accommodation, the results are very similar. Some of the experiments employed a fixation point; others allowed the observers to move their eyes freely. When disparities are small, the results are again very similar.

# RELATIVE DISTANCE TASKS

I will describe performance on two classes of distance tasks. The first are called relative distance tasks; they are tasks in which an observer adjusts the position of light points by remote control until they satisfy some relative distance criterion (Foley, 1978, 1980). Examples of such criteria are shown in figure 2. In each case the view is from above; the oval represents the observer's head and the dots represent stimulus lights. In the apparent fronto-parallel plane (AFPP) task, one point of light is fixed and the observer moves other lights so that they appear to lie in the vertical plane through the fixed light that is parallel to the vertical plane through the eyes or, in other words, a plane that is perpendicular to straight ahead. The apparent equidistant circle (AEDC) task is very similar, except that the lights are set so that they are perceived to lie on a circle with the observer at the center. In the apparent distance bisection (ADB) task, one point is fixed and the observer adjusts a second point so that the distance between the two points is perceived to equal the distance from the observer to the near point.

Typical performances in these tasks are illustrated in the second row for three distances of the fixed point. In each task there is one distance at which the physical configuration corresponds to the perceived configuration. This distance is generally within the range of 1-4 m. At other distances, there are systematic errors in the settings. At far distances, variable points are set too far, and at near distances, they are set too near, relative to accurate performance. Although there are individual differences in the magnitude of the errors, errors of this kind are reliably found. (For many observers, one side of the configuration is set closer than the other (skewing). This can be accounted for by a very small difference in magnification in the two eyes. This is incorporated in a general theory of binocular distance perception (Foley, 1980), but it is not considered in this article.)

I propose that these errors can be explained by the misperception of the egocentric distance to the fixation point, or, in the absence of a fixation point, to a reference point that depends on the configuration of points. To test this idea we must consider how the pattern of disparities produced by the observer compares with the pattern of disparities corresponding to the physical configuration specified by the instructions. By pattern of disparities I mean the function that relates binocular disparity to direction. The left side of figure 3 shows this function for physically fronto-parallel planes (PFPP) at different distances and the right side shows the same function for AFPP at different distances. If all the error in the AFPP settings is due to the misperception of the distance to the fixation point, then the function for an AFPP should be identical to the function for a PFPP, but generally this will be a PFPP at another distance. This is what the experiments show. For example, an AFPP at 1.2 m has less disparity than a PFPP at 1.2 m, but corresponds to the same disparity pattern as a PFPP at 1.45 m. Patterns of disparities obtained in the AEDC task also correspond closely with disparities produced by physically EDCs at other distances. Thus, the experimental settings can be accounted for by the hypothesis that the observer misperceives the egocentric distance to the configuration and produces the pattern of disparities appropriate to the misperceived distance.

This hypothesis has several important implications. First, the fact that the pattern of disparities changes with the distance to the fixed point implies that there is an egocentric distance signal related to the vergence of the eyes, and this egocentric distance signal is not accurate. Second, effective binocular disparity equals binocular disparity. This is illustrated in figure 1. In general, the distance to point $r$ will be misperceived. But if $r$ is misperceived, any other point $i$ will also

be misperceived, so that the difference between the effective binocular parallaxes equals the difference between the binocular parallaxes. I call this the effective disparity invariance principle.

The data from relative distance tasks may be used to infer the perceived distance to the fixation point or to the reference point. The simplest way to conceptualize this is t o imagine a more complete set of functions on both sides of figure 3. Then, for each pattern on the right, we find the matching pattern on the left. The distance on the right is the physical distance that corresponds to the perceived distance on the left. This perceived distance is a concave downward function of physical distance, as is shown by the solid line on the left side of figure 4. When both physical distance and perceived distance are transformed to parallaxes, their relation becomes linear, as is shown by the solid line on the right side of this figure. I call the curved function on the left the reference distance function and the linear function on the right the reference parallax function.

## EGOCENTRIC DISTANCE TASKS

Next consider a different class of tasks—egocentric distance tasks. An egocentric distance task is one in which an observer indicates the distance from herself or himself to visual targets (Foley, 1977, 1985). Several different indicators have been used, but I have relied on two, verbal reports of perceived distance and pointing with an unseen hand. In the pointing experiments a horizontal board just beneath the targets prevents the observer from seeing his or her hand or arm. I will describe two simple experiments.

In the first experiment the stimulus is a single light point in dark surroundings. It is straight ahead. Pointed distances and reported distances from such experiments are shown in figure 4. The smooth curves shown have parameters that are close to the average values fitted to the data of five observers (Foley, 1977). On the left, indicated distance is plotted against physical distance, and on the right, the same values are plotted as binocular parallaxes. The functions on the left have the same form as the reference distance function; those on the right, the same form as the reference parallax functions.

But there is a complication: Verbal and manual indicators do not agree, and neither, in general, agrees with the function inferred from the relative distance tasks, which tends to lie between the verbal and manual functions. Since the indicators do not agree, both cannot correspond to perceived distance. I have defined perceived distance as the distance inferred from the relative distance tasks. When expressed as parallaxes, this value and the values indicated by pointing and verbal reports are all linearly related. This means that egocentric distance tasks can be used to test the implications of the theory. It is very important, however, to distinguish between perceived distance and indications of it. In figure 4 only the solid lines derived from the relative distance tasks correspond to perceived distance and reference parallax; the other lines describe indicated distance and indicated parallax.

When the eyes move freely, there is one point the perceived distance of which is given by the reference distance function. I call this point the reference point. Perceived distances of all other points are determined by their disparities relative to this point. There are several ways to determine the reference point. The most obvious is to measure the effective parallax of each point in the configuration and then determine how these are related to the reference parallax function. This analysis has been carried out only for the case of two-point configurations (Foley, 1985). Here the

parallax of the reference point is a weighted average of the parallaxes of the points, with the farther point tending to receive the greater weight. Thus the reference point need not correspond to any point of the configuration, although sometimes it may.

## DISCUSSION

Figure 5 is a schematic diagram illustrating the process of binocular distance perception. The visual system generates both binocular parallax and binocular disparity signals in response to the optic array. The binocular parallax signals determine a single reference point and its corresponding value of effective binocular parallax. Here this is shown as an outflow from an eye movement control center. For each point $i$, the disparity of $i$ relative to the reference point is added to the effective reference parallax to give the effective parallax of the point. This value undergoes an indicator-specific linear transform to yield the indicated binocular parallax, which, in turn, determines the response.

When multiple cues are present, including perspective cues, distance perception is more accurate; however, the evidence indicates that there are systematic errors in distance perception under most cue conditions. There are several studies that have examined apparent distance bisection under such conditions. Although results have varied widely, no study has found consistently accurate bisection over a wide range of distances. The most common result is that the farther interval is set larger than the nearer one. There are also several studies that have obtained verbal reports of perceived distance under multiple cue conditions. The data are often fitted with a power function and the power is generally less than 1. An experiment limited to distances less than 70 cm yielded an accelerating verbal report function and a decelerating pointing response function (Foley, 1977). When the inverse output transforms derived from binocular experiments are applied to these data, both verbal and manual responses yield the same parallax function with a slope of about 0.8. The conclusion is that distance perception is generally inaccurate, even in the presence of multiple cues.

How can we perform accurately with respect to distance when distance perception is inaccurate? I can only answer this speculatively because the experiments needed to answer it scientifically have not been done. I hypothesize that we learn to behave accurately on the basis of feedback. This learning cannot be once and for all because the errors that it compensates for vary continuously with changing cue conditions. I hypothesize that the output transforms that I have proposed to explain open-loop performance are modified by feedback to compensate for perceptual errors.

What implications does this have for the design of visual displays? I would expect that most visual displays evoke erroneous distance percepts. I expect this because even a three-dimensional scene with multiple cues evokes erroneous percepts, and most displays both eliminate cues and introduce cue conflicts, both of which are associated with increasing errors. In principle, it might be possible to create a display that would evoke accurate percepts, at least in some limited domain, but I doubt the wisdom of attempting this. The perceptual-motor system is designed to make rapid compensation for certain forms of error, especially those that can be described by linear transforms of the reference parallax function. Displays that produce errors of this form should suffice to direct behavior. But every time a display is used to direct behavior in the real three-dimensional space,

performance with feedback is necessary to calibrate the output transforms, just as performance with feedback is necessary when a three-dimensional scene directs behavior.

## REFERENCES

Foley, J. M.: Effect of Distance Information and Range on Two Indices of Visually Perceived Distance. Perception, vol. 6, 1977, pp. 449-460.

Foley, J. M.: Primary Distance Perception. In Handbook of Sensory Physiology, R. Held, H. Leibowitz, and L. Teuber, eds., Berlin: Springer-Verlag, vol. 8, pp. 181-213.

Foley, J. M.: Binocular Distance Perception. Psych. Rev., vol. 87, 1980, pp. 411-434.

Foley, J. M.: Binocular Distance Perception: Egocentric Distance Tasks. J. Exp. Psych.: Human Percept. Perform., vol. 11, 1985, pp. 133-149.
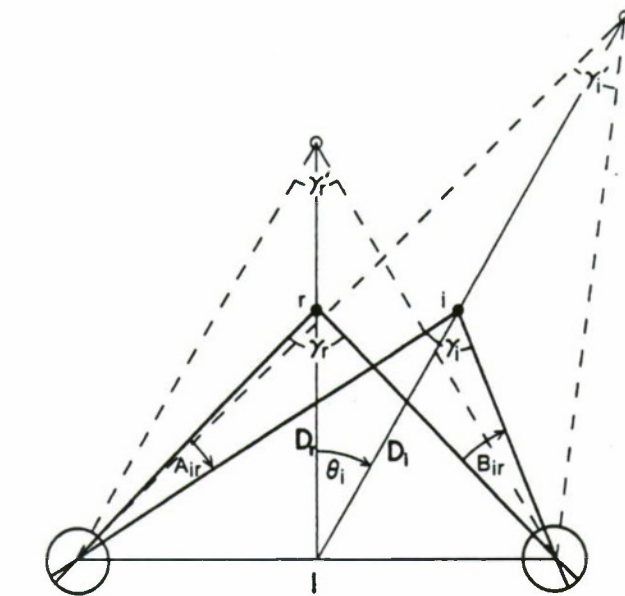
**Variables**

| Binocular Parallax | Binocular Disparity |
|---|---|
| $\gamma_i = (I/D_i)\cos\theta$ | $\Gamma_{ir} = \gamma_i - \gamma_r$ |
| Effective Binocular Parallax | Effective Binocular |
| $\gamma'_i = (I/D'_i)\cos\theta$ | $\Gamma'_{ir} = \gamma'_i - \gamma'_r$ |

Figure 1.- Variables used in this article. The figure is a top view of the horizontal eye-level plane. The large circles at the bottom represent the two eyes and the solid dots labeled $r$ and $i$ correspond to two stimulus points. The expressions at the bottom of the figure define the four variables. $I$ is interocular distance. $D$ is radial distance to a point. $\theta_i$ is horizontal direction of a point relative to straight ahead. $D'$ is perceived radial distance.
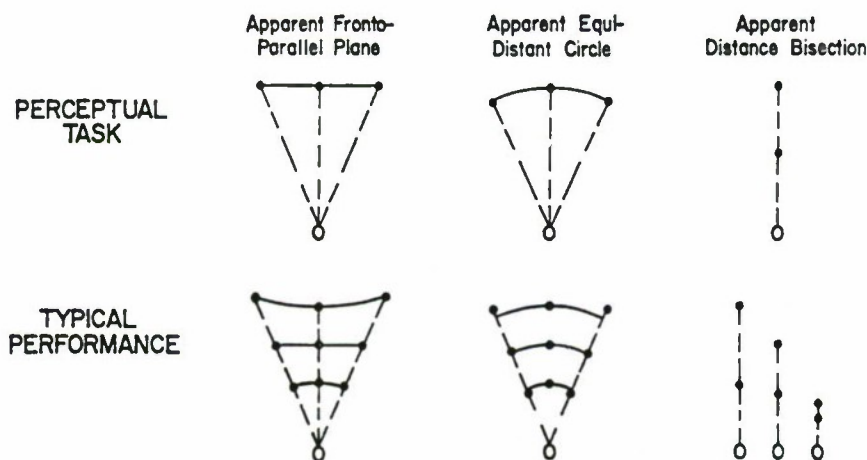


Figure 2.- Illustration of three relative distance tasks (top) and typical performance for observers who show no skewing (bottom). The physical configuration corresponds to the perceptual criterion only at one distance, which is typically between 1 and 4 m. The diagram is not to scale.

3-6
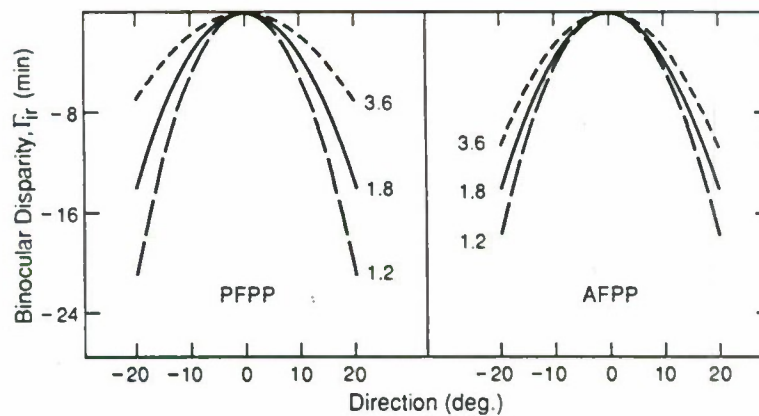
Figure 3.- Binocular disparity as a function of horizontal directions for PFPP and AFPP; the smooth curves describe the results of a typical observer. Each function is shown for three distances of the fixed center point: 1.2, 1.8, and 3.6 m. For this observer the functions correspond at 1.8 m. As distance becomes greater or less than this, the disparities that correspond to the AFPP change less than those corresponding to a PFPP.
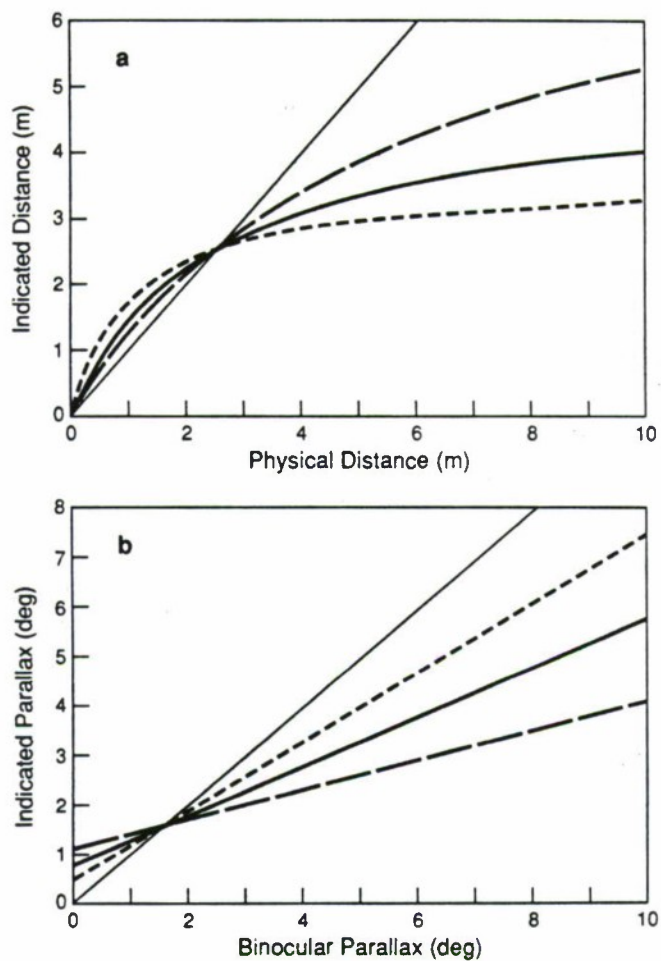
Figure 4.- a) Perceived (or indicated) distance as a function of target distance. Perceived distance inferred from relative distance tasks —; perceived distance indicated by manual pointing -•-; perceived distance indicated by verbal report - - -. b) The same three functions expressed as parallaxes.
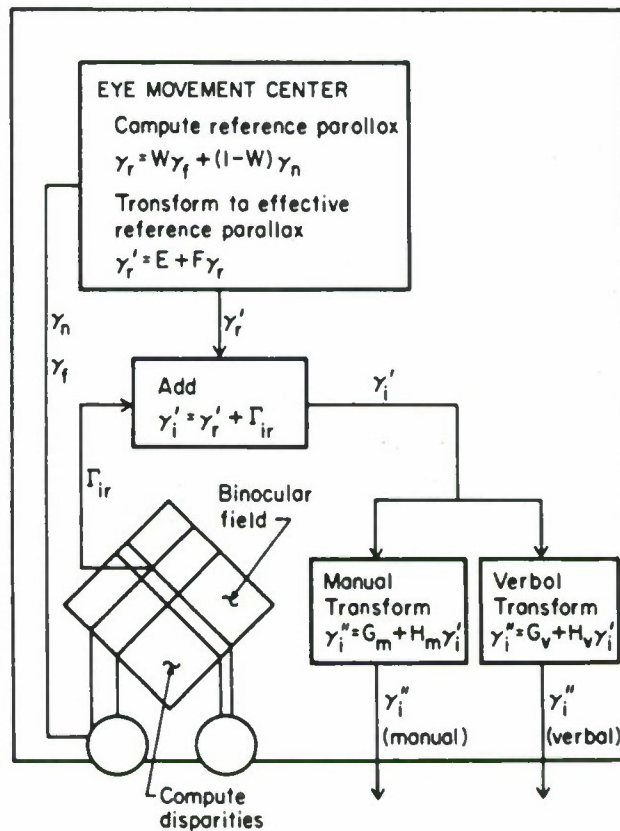
Figure 5.- Diagram summarizing the formal operations of the model in a way that suggests under-lying structures and processes (from Foley, 1985).

# PARADOXICAL MONOCULAR STEREOPSIS AND PERSPECTIVE VERGENCE

J. T. Enright
Scripps Institution of Oceanography
La Jolla, California

## SUMMARY

The question of how to convey depth most effectively in a picture is a multifaceted problem, both because of potential limitations of the chosen medium (stereopsis? image motion?), and because "effectiveness" can be defined in various ways. Practical applications usually focus on "information transfer," i.e., effective techniques for evoking recognition of implied depth relationships, but this issue depends on subjective judgments which are difficult to scale when stimuli are above threshold. Two new approaches to this question are proposed here which are based on alternative criteria for effectiveness.

Paradoxical monocular stereopsis is a remarkably compelling impression of depth which is evoked during one-eyed viewing of only certain illustrations; it can be unequivocally recognized because the feeling of depth collapses when one shifts to binocular viewing. An exploration of the stimulus properties which are effective for this phenomenon may contribute useful answers for the more general perceptual problem.

Perspective vergence is an eye-movement response associated with changes of fixation point within a picture which implies depth; it also arises only during monocular viewing. The response is directionally "appropriate" (i.e., apparently nearer objects evoke convergence, and vice versa), but the magnitude of the response can be altered consistently by making relatively minor changes in the illustration. The cross-subject agreement in changes of response magnitude would permit systematic exploration to determine which stimulus configurations are most effective in evoking perspective vergence, with quantitative answers based upon this involuntary reflex. It may well be that "most effective" pictures in this context will embody features which would increase "effectiveness" of pictures in a more general sense.

## INTRODUCTION

One of the central issues involved in spatial display is the question, "What is the most effective way to convey three-dimensional depth in a pictorial representation?" This article deals only with a very restricted approach to that question, being confined to representations without stereopsis and without image motion; and so the problem addressed here should probably be rephrased, "What is the *third* most effective way of conveying depth in pictures?" Such rephrasing seems appropriate because there can be little doubt that the most effective representations of the third dimension are those which involve stereopsis; and that the second most effective way to convey a feeling for depth is through use of image motion: optical flow patterns, image shear, motion parallax and the like. When both stereopsis and image motion are excluded, one is dealing with no more than third best; and the rephrased question is in some ways like asking what is the best way to participate in a footrace, subject to the precondition that the runner's feet be tied together by his shoelaces.

Nevertheless, the question of how best to convey the third dimension in a static pictorial representation has been of central concern to artists for many hundreds of years; and the result of that interest is an organized body of technique, collectively known as perspective, to deal empirically with that problem. One might well ask, then, whether there is any hope for deriving new answers to this question—if thousands of artists, throughout their careers, have been experimenting for centuries with just this objective in mind. The honest reply is that this article has no new answers to offer, no new tricks to suggest. Instead, it focuses upon two interesting phenomena involving the perception of and response to depth in illustrations—phenomena which seem to me to have the potential of providing more quantitative answers to the question, "How can depth be more effectively represented?" These phenomena suggest research programs for the future, which would address this question within certain restricted contexts, and it is conceivable that the answers might be applicable to other, more general contexts as well. The hope is that such research might provide general, quantitative rules for optimizing the depth impression which is conveyed by the stimulus field in an illustration.

## PARADOXICAL MONOCULAR STEREOPSIS

The first of the phenomena of interest here is a remarkable and relatively little-known sort of depth perception which was described by the French visual scientist, Claparède, in a brief article published in 1904; he christened this visual experience "paradoxical monocular stereopsis." The essence of Claparède's message is that if certain pictures which illustrate a three-dimensional scene—drawings, paintings or photographs—are carefully examined *with one eye covered*, a truly compelling sense of depth can sometimes be obtained, an effect nearly as striking as looking into a stereoscope. Once this sort of perception has been achieved, it can be sustained while continuing to inspect the picture, and one might suspect that it results simply from thinking about and focusing attention on the illustrated subject matter. It is easy to demonstrate, however, that something unusual is involved, because the moment that the other eye is opened, to see the picture binocularly, the anomalous 3-D effect vanishes; the picture flattens out just as suddenly and completely as when one *closes* one eye while looking into a stereoscope.

High-quality, well-printed color photographs of outdoor scenes, of the sort found in magazines like *National Geographic* and *Arizona Highways*, often provide good material for demonstrating this sort of depth perception, but one of the most interesting aspects of paradoxical monocular stereopsis is how difficult it is to predict whether a given illustration will be effective in evoking the response. The compelling impression of depth is not simply a response to monocular viewing of all illustrations which show a three-dimensional scene, but to certain configurations of stimuli. The question therefore arises, "What is the most effective way to evoke paradoxical monocular stereopsis with an illustration?" This is, of course, a much more limited question than asking what is the most effective way to convey depth in a picture, but it may be more tractable. One has available the clear-cut criterion, "Does the (supplementary) depth impression flatten out, when switching over to binocular viewing?" Furthermore, although the best stimuli for paradoxical monocular stereopsis may not turn out to be fully congruent with the stimuli which are optimal for conveying a three-dimensional impression during binocular viewing, preliminary evidence suggests that if a picture is effective in evoking paradoxical stereopsis, it will at least give a satisfying and convincing impression of depth during binocular viewing.

A search of the published literature indicates that there have apparently been no systematic investigations of which kinds of pictures best evoke paradoxical stereopsis; and in fact, I have encountered less than a dozen references, in the entire 80-year interval since Claparède's (1904)

initial description of the phenomenon, in which this sort of depth perception is even mentioned (e.g., Pirenne, 1970; Schlosberg, 1941; Ames, 1925; Streigg, 1923; and the references cited there). Qualitative preliminary testing indicates that there is good agreement among subjects, in the sense that certain pictures seem to be very effective stimuli for everyone, so the project of exploring stimulus optimization should be relatively easy to carry through, with a relatively modest number of subjects. And if the illustrations which are to be used were to be carefully selected, it seems very likely that an organized body of rules will emerge which characterize the optimal stimuli.

## PERSPECTIVE VERGENCE

In the brief article in which Claparède (1904) described this unusual sort of depth perception, he also proposed an interesting hypothesis about the mechanisms responsible. He speculated that during monocular inspection of a picture, the covered eye would be free to make vergence movements which might correspond to the relative distances implied by the illustration (converging, then, for apparently near objects and diverging for more remote ones), just as changes in vergence accompany binocular inspection of a real, three-dimensional scene. He pointed out that vergence changes of this sort could not take place during binocular viewing of a picture because of the demand for fusion; and he further proposed that this sort of postulated vergence movement might be responsible for the compelling sense of depth evoked during monocular viewing. Apparently there has been no test of Claparède's hypothesis, nor even any restatement of it, in the subsequent 80 years; a recently initiated research program, however, has provided compelling evidence that Claparède was essentially correct in his speculation about eye movements (Enright, 1987a; Enright, 1987b). Vergence changes of the sort he postulated do, indeed, take place when inspecting a picture of a three-dimensional scene with one eye covered—though whether those eye movements are responsible for paradoxical stereopsis remains an open question, and one which will be much more difficult to investigate.

## METHODS

The experimental equipment which was used in this eye-movement research is extremely simple, both in principle and in practice (Fig. 1). The subject sits with head held firmly in place by a bite board and headrest while two video cameras monitor eye position from somewhat below the line of sight. The output of the cameras is combined with an image splitter and recorded for subsequent analysis; the sum of the two distances between iris margins and the image-splitting line is an index for vergence state. The illustrations to be viewed are mounted at about 30 cm from the subject's eyes, and an obstruction is placed a few centimeters in front of the nondominant eye, at a level which hides the picture from that eye, but permits the camera to record eye position. While viewing the picture monocularly, the subject changes fixation at intervals of 2 to 3 sec, between points which are at different implied distances away. Single-measurement precision of the recording method is about 6 arcmin for each evaluation of eye position, and averaging results over repeated tests can further reduce the influence of random measurement error; but the between-trial variability within a given test session for a given subject and target is sufficiently large that a more precise monitoring technique could not appreciably improve the reliability of the estimates of average response; the variability in the eye movements from one refixation to the next limits precision of the estimates, as reflected in the standard errors.

# RESULTS

An excerpt from a longer recording is shown in Fig. 2, made while a subject changed fixation from the upper front corner to the upper back corner of the perspective drawing of a small box (target illustrated in Fig. 3). Concurrent with the recording, a three-position switch, which was connected to two tone generators, was activated by the subject to indicate the fixation point; the timing of those signals is shown as open and solid bars in Fig. 2. It is, then, quite clear that convergence occurred while fixating on the apparently nearer corner of the box, and divergence while fixating on the farther corner. A simple summary value for the typical vergence-change response can be obtained from such a recording, based on measuring one value of vergence state for each steady-state fixation, and then calculating differences between successive values; in this case, the average change in vergence, over 20 fixations, was 68 arcmin ± 8 arcmin. In Fig. 3, this summary value is shown for Subject 1, along with five other values for her, each with this same target, each recorded on a different day; and values of average vergence change are also shown there for another eight subjects with this target. Average vergence change, based on the method of calculation, could in principle also be negative (i.e., contrary to the perspective implication of the drawing); in fact, however, all 24 measured values are positive, and all except one of the results are statistically significant, most of them at the 0.01 level. In other words, the subjects all showed consistent vergence changes during changes in fixation point in this drawing; and those vergence changes corresponded in direction with the relative distances implied by the perspective of the drawing. For those who may be concerned about the reliability of this simple and unconventional method of recording eye movements, it is worth mentioning that the basic result of Fig. 3 has now been replicated for other subjects in two other laboratories, each of them using a fundamentally different and more familiar measurement technique. I have proposed (Enright, 1987a) that these oculomotor responses to pictorial representations be called "perspective vergence."

Before considering additional details of the responses which have been measured for other kinds of illustrations, it seems worthwhile to try to place perspective-vergence responses into some sort of broader context. A phenomenon which is now called "proximal vergence" has long been known to visual physiologists, an eye-movement response which has been attributed to "knowledge of nearness" (Maddox, 1893). Although vergence responses to perspective representations have not been previously studied, it is probably appropriate to consider perspective vergence to be a subcategory of "proximal vergence" (Hokoda and Ciuffreda, 1983). It is important, however, to distinguish between these responses and another subcategory known as "voluntary vergence": some trained subjects can cross or uncross their eyes at will, even in total darkness. Many lines of evidence indicate, however, that the eye-movement responses to perspective illustrations are instead the result of an involuntary reflex. It is conceivable—even likely—that training or an "act of will" might enhance the responses, but fully naive, untrained subjects also show comparable behavior in their first test session—even subjects who are fully unaware that convergence is the appropriate response to objects which are nearby. They show this response even though they are uninformed about the purpose of the experiment, even though they have no visual feedback or other clues to tell them whether vergence has changed—much less whether the response was "as intended." Perspective vergence is an automatic response to components of the visual stimulus field—truly a reflex. Furthermore, at least certain components of the stimulus field which evoke this kind of response are apparently not a reflection of learning or prior experience, but instead represent built-in constraints on the visual system—although it seems likely that "learning" may also play a role—that prior visual experience with our three-dimensional world may build upon and supplement those components which are "hard-wired" into the system. Because of the reflex nature of the responses, an evaluation of illustrations, in terms of the magnitude of the

vergence responses evoked, represents something far more substantial than can be achieved by asking for subjective opinions about picture quality.

An experimental program has been initiated, designed to determine what features of an illustration enhance or inhibit this oculomotor response. The results of Fig. 4 summarize some of the kinds of data which have been obtained, with modest variations on the compositional theme of a single rectangular box. Despite the large inter-subject differences in response magnitude for a given picture, as shown in Fig. 3, there are remarkably consistent cross-subject *changes* in response magnitude for particular alterations in the picture; hence, the ratio of response for a given picture to the same subject's response for a standard, represents a reliable way of demonstrating the relative effectiveness of various representations in evoking perspective vergence. Doubling the size of the picture in all dimensions, for example, reliably led to an increase of about 50% in response magnitude (Fig. 4 vs. Fig. 4B); inverting the picture led to a reduction in response (Fig. 4A vs. Fig. 4C), with 7 of 9 subjects showing smaller vergence changes. A reduction in the inclination of the box (with only minor other modifications in line spacing) led to a drastic reduction in response magnitude (Fig. 4B vs. Fig. 4D); for 8 of the 9 subjects, the response was even smaller than that to the "standard" picture, which shows a box half the size (Fig. 4A). When a cross-hatched lid was superimposed upon a box which was in the relatively ineffective orientation, response magnitude increased for all 9 subjects (Fig. 4D vs. Fig. 4E), but when a similar lid was superimposed on a box with more effective orientation, it tended to reduce the response (Fig. 4A vs. Fig. 4F; 8 subjects out of 9). In all cases, there was remarkably good cross-subject agreement in the way in which a given change in the drawing affected magnitude of the response (details in Enright, 1987a).

One other closely related kind of target has been tested, which is not shown in this figure; three-dimensional cardboard models of the boxes shown in Figs. 4A and 4D were constructed and photographed from 30 cm with illumination which produced a distribution of light and shadow, and prints of those photos, at appropriate scaling, were tested as targets. The rationale for this approach is that shading might enhance the resulting vergence changes. In these tests there was indeed a slight but significant increase in response for the box shown with suboptimal orientation (Fig. 4D), but no significant change—in fact a slight decrease—for the more optimally oriented box (Fig. 4A).

The vergence responses of this same group of 9 subjects have also been tested with a set of more complex pictorial representations: photographs which reproduce five classical paintings and an etching; and those experimental results have offered further hints about the kinds of stimuli which can be effective in evoking perspective vergence. By using a portrait by Rembrandt, for example, statistically significant vergence changes in the appropriate direction (nearly as large as those for the "small-box" drawing [Fig. 3]), were evoked in all 9 subjects by a change in fixation from the nose to the ear of the portrayed philosopher and back again, although no suggestion of *linear* perspective was evident in the picture, and the implied difference in distance between the fixation points was quite small (ca. 10 cm, at a distance of 2 to 3 m from the viewer). One landscape scene evoked strong responses in every subject tested, and another outdoor scene, in which linear perspective was conspicuous, did not lead to statistically significant results for *any* of the subjects. Again, then, there was very good cross-subject agreement, in terms of which artworks were effective stimuli and which were not.

# DISCUSSION

The cross-subject consistency in terms of response magnitude demonstrates that in measuring perspective vergence we are dealing with relatively general characteristics of the oculomotor response system; but the experiments conducted so far do no more than define a few of the dimensions of the multidimensional coordinate system implied in the question, "What is the optimal stimulus for this response?" There seems to be clear non-additivity (a cross-hatched surface between fixation points enhances a response, or it does not, depending on context), which considerably complicates the exploration of these dimensions. Furthermore, it is by no means clear that the rules which might be derived from a line drawing of a cubical box can be generalized to other sorts of figures; nor do the available data define an optimum point in any stimulus dimension. Consider, for example, the conspicuous effect of tilt of the opening on responsiveness (Fig. 4B vs. 4D): while it seems clear that a 22° tilt (4B) is much more effective than an 11° tilt (4D), there is presumably a continuous function relating responsiveness to inclination in the illustrated box, with a maximum someplace between 0° and 90°; and it may well be that 22° is far removed from that optimum tilt. The necessary experiments to explore this dimension should be enlightening— but the existence of nonlinearities cautions against overgeneralization.

The consistently positive responses to the Rembrandt portrait demonstrate that the dimensions which must be explored in any complete attempt to define optimal stimuli go far beyond the systems of lines and angles which constitute linear perspective. The opportunity to explore the question of stimulus optimization offers exciting promise for the future, but it is self-evident that the available data do not even adequately define the dimensions of the problem. Beyond the issue of stimulus optimization, the intriguing possibility exists that perspective vergence responses may provide an objective metric for evaluating the general effectiveness of an attempt to convey depth in a picture: that oculomotor responsiveness may prove to be well correlated with subjective perceptual responsiveness to pictorial implications of depth. Such a correlation would be a necessary— but not a sufficient—condition for establishing the validity of Claparède's most interesting speculation: that perhaps vergence movement itself contributes to the perception of paradoxical monocular stereopsis.

## ACKNOWLEDGMENT

# REFERENCES

Ames, A., Jr. (1925): The illusion of depth from single pictures. J. Opt. Soc. Amer., 10, 137-148.

Claparède, E. (1904): Steréréoscopie monoculaire paradoxale. Annales d' Oculistique, 132, 465-466.

Enright, J. T. (1987a): Perspective vergence: oculomotor responses to line drawings. Vision Res., 27, 1513-1526.

Enright, J. T (1987b): Art and the oculomotor system: perspective illustrations evoke vergence changes. Perception, 16, 731-746.

Hodoka, S. C. and Ciuffred, K. J. (1983): Theoretical and clinical importance of proximal vergence and accommodation. In Vergence Eye Movements: Basic and Clinical Aspects. (Ed. by C. M. Schor and K. J. Ciuffreda) pp. 75-97. Butterworths, London.

Maddox, E. E. (1893): The Clinical Use of Prisms: and the Decentering of Lenses. Second Edition. Bristol, John Wright and Sons.

Pirenne, M. H. (1970): Optics, Painting and Photography, p. 95. Cambridge Univ. Press.

Schlosberg, H. (1941): Stereoscopic depth from single pictures, Amer. J. Psychol., 54, 601-605.

Streiff, J. (1923): Die binoculare Verflachung von Bildern, ein vielseitig bedeutsames Sehproblem. Klin. Monatsbl. Augenheilkunde, 70, 1-17.
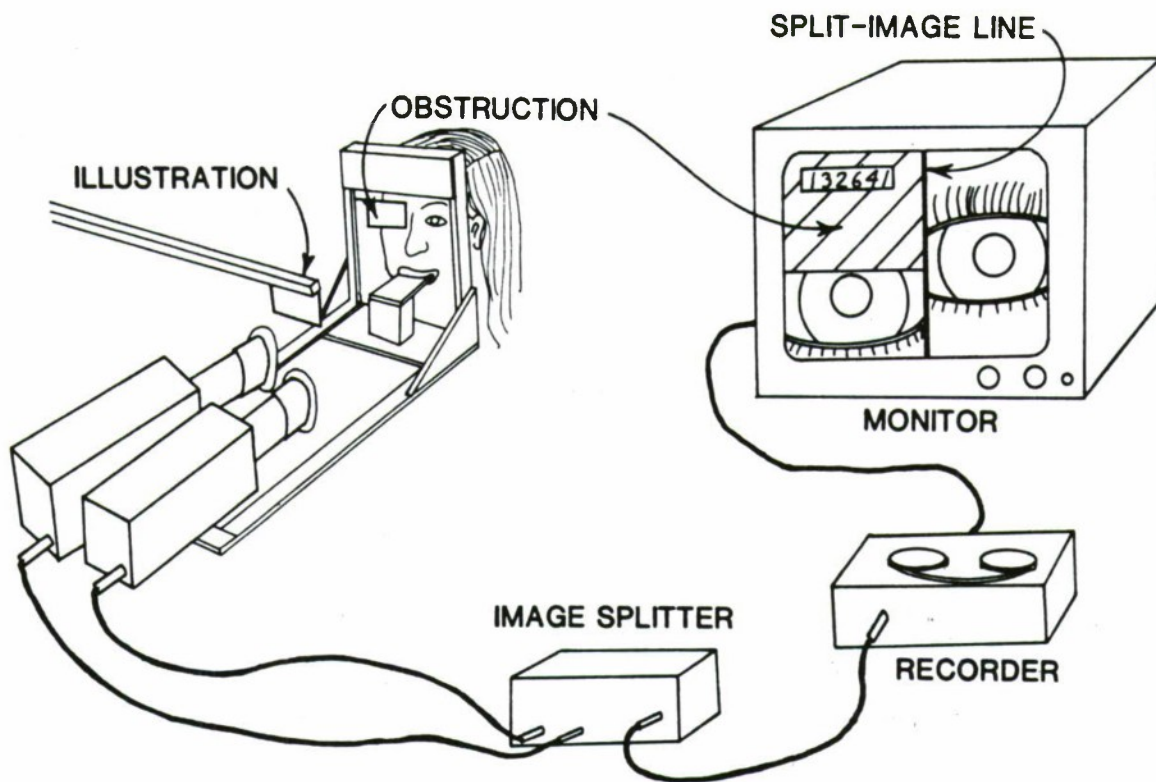
Figure 1.– Diagram of the equipment and setup used for recording eye position while viewing illustrations.
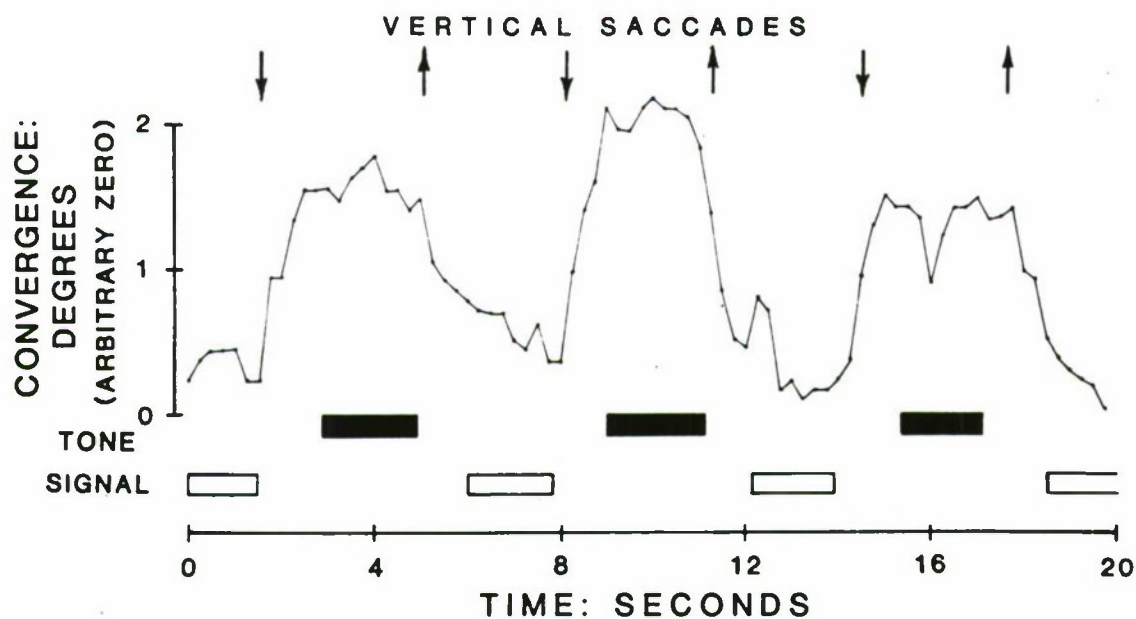


Figure 2.– Excerpt from a recording made while Subject 1 alternated monocular fixation between apparently nearer and apparently farther topside corners in a line drawing of a small cubical box (picture shown in Fig. 3 and as "Standard" in Figure 4). Bars beneath graph correspond to the timing of tone signals; solid bars represent fixation on "near" corner, open bars represent fixation on "far" corner. (Reprinted with permission from Vision Res. 27, J. T. Enright, "Perspective Vergence: oculomotor response to line drawings," Copyright 1987, Pergamon Journals Ltd.)
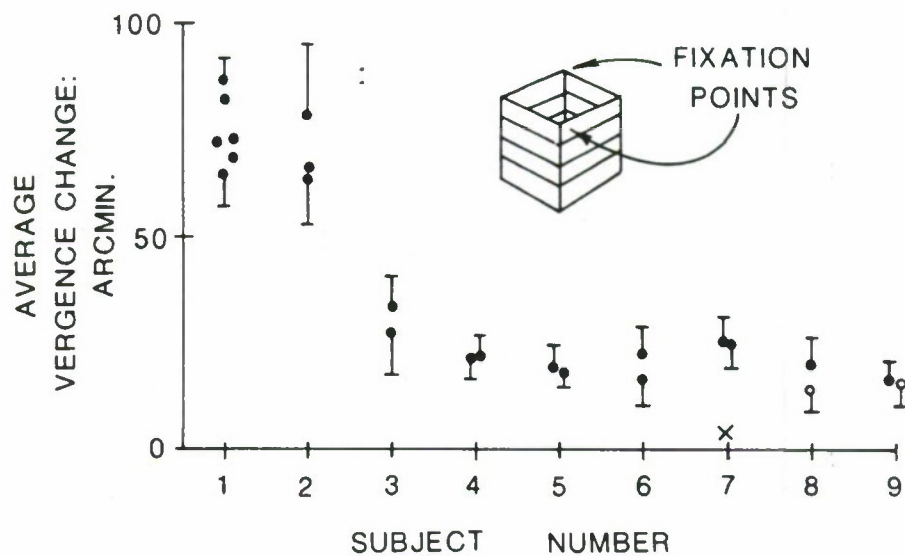
Figure 3.– Summary of average vergence changes made by 9 subjects in conjunction with changes in fixation on the line drawing of a small cubical box; each point represents average value during a separate test session, with standard errors based on N of 10 (20 changes in fixation).
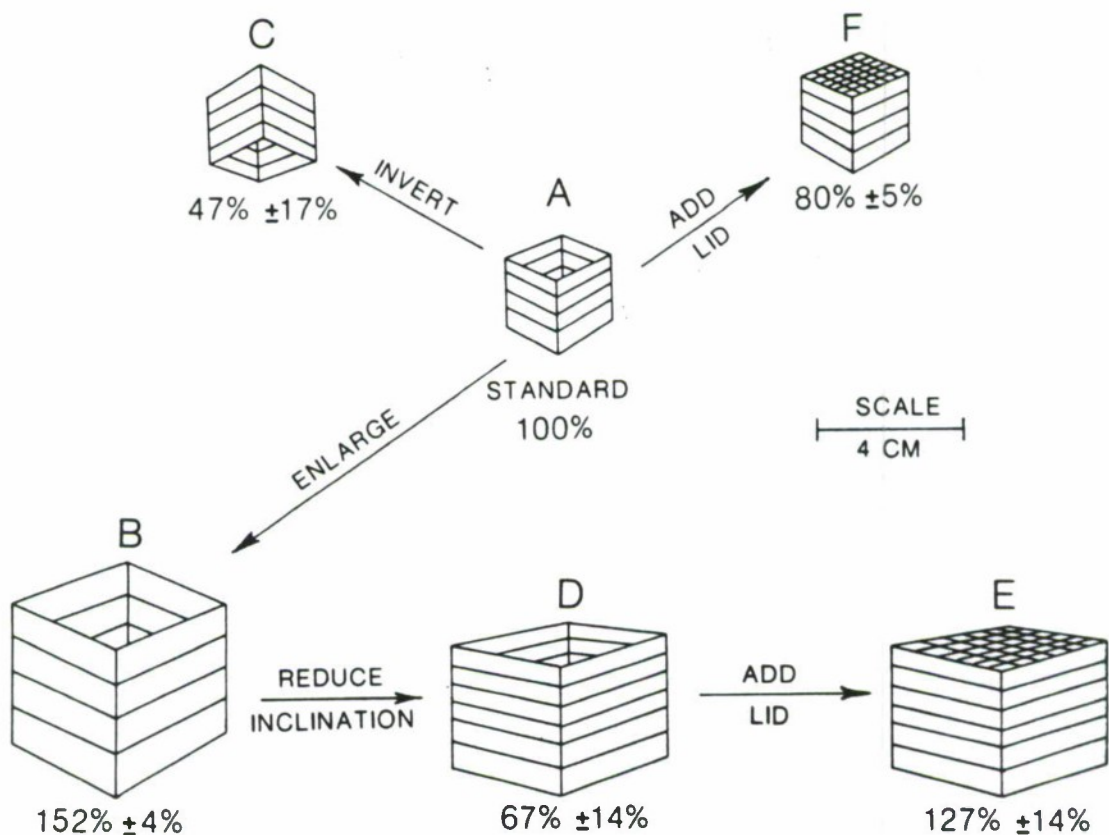


Figure 4.– Cross-subject values, and their standard errors, for 100 times the ratio: "average vergence change for a given drawing," divided by the same-subject value of "average vergence change for 'standard' illustration."   N = 3 for part B, N = 9 for all other parts.

# SPATIAL PERCEPTION: OTHER CUES

# SEEING BY EXPLORING

Richard L. Gregory
Department of Psychology
University of Bristol
Bristol, England

The classical notion of how we see things is that perception is passive—that the eyes are windows, and in floods reality. This was how the Greeks saw perception, and it is the basis of the accounts of the seventeenth and eighteenth century Empiricist philosophers. But physiological work of the nineteenth century cast doubt on this view that perception is passive acceptance of reality. The doubt arose from discoveries of elaborate neural mechanisms, of the delay of signals, and of the time required to process the signals and then make decisions. The doubt was fueled by interest in phenomena of visual and other illusions; for how could passively accepted truth be illusory? It was clear to Hermann von Helmholtz and others and hundred years ago that illusions suggest active processes of perception, which do not always work quite correctly or appropriately. This discovery, and surely this was an important discovery, was not all popular with philosophers—for perception as the principal basis for true statements became suspect. Worse, evidently perception needed scientific backup (and indeed, what was discovered with instruments did not always agree with how things seem to the senses), so philosophers lost out to scientists as the discoverers and arbiters of truth. Fortunately for them, scientists often disagree on their observations, and how they should be interpreted, so philosophy gradually took on other roles, especially advising scientists what to do.

Perhaps curiously, perception is not at the present time a popular topic for philosophers. This must be partly because scientific accounts of perception have now gone a long way away from appearances. They depend on physiological and psycho-physical experiments (as well as curious phenomena including various kinds of illusions) which require technical investigation and do not fall within traditional concepts of philosophy. For example, it has become clear over the last 20 years or so that visual perception works by selecting various features from the environment, by specialized information channels of the eye and brain. This is an extension of the nineteenth century physiological concept of the Specific Energies of nerves, suggested by the founder of modern physiology Johannes Muller (1801-58). His notion that there are many special receptors and neural pathways, each giving its own distinct sensation, has recently been confirmed and extended for touch, hot and cold, and tickle (Iggo, 1982). In vision, various features (such as the position and orientation of edges, direction and velocity of movement, stereoscopic depth, brightness, and colors) are signaled by dedicated channels having special characteristics for transmitting and analyzing significant features of the world. There are also "spatial frequency" channels, tuned to separations of features, which suggest that spectral analysis plays some part in pattern recognition. All this implies that a great deal of parallel processing goes on in the visual system—leading to integrated pattern vision in which many sources of information, sensory and stored from the past, come together—to give powerfully predictive hypotheses, which are our reality of the object world. It seems appropriate and useful to think of perceptions as "hypotheses" (Perceptual Hypotheses) by analogy with the hypotheses of science which make effective use of limited data for control and prediction (Gregory, 1974, 1981).

We may go on to ask further what, perceptually, is an object? What is accepted or seen as an object depends greatly on use—on what is handled, or what behaves, as a unit. It seems that we

map the world into individual objects in infancy, by exploring with our hands and discovering what can be pushed or pulled as units, and generally how things behave to us and to each other. Thus when we read a book, each page is an object, as we turn them separately; but on the shelf each book is an object, as they are selected and picked as a unit. And on a printed page letters, words, sentences, or paragraphs may be units, according to how we read. Perceptual units are set up early in life, but it is an interesting possibility that new structuring might be continued throughout adult life—by continuing to explore the world with our hands and eyes. Then we might continue the remarkable perceptual and intellectual development of childhood throughout life. This is the hope (one might almost say religion) of interactive "hands-on" science centers, including the *Exploratorium* founded by Frank Oppenheimer in San Francisco, and the *Exploratory* we have started in Bristol (Gregory, 1986). They allow people of all ages to discover the world of objects (and something of science and technology, as well as their own perceptions) by active exploration.

The importance of experience through interaction with objects was impressed upon me 25 years ago when my colleague Jean Wallace and I studied the rare case of someone (S. B.) who, after being effectively blind from infancy, received corneal grafts in middle life. This is the situation envisaged by John Locke, following a letter he received from his friend Samuel Molyneux who asked, "Suppose a man born blind, and now adult, and taught by his touch to distinguish between a cube and a sphere of the same metal. . . . Could he distinguish and tell which was the globe, which the cube?" Locke (1690, Bk. II, Chapt. 9, Sect. 8) was of the opinion that "the blind man, at first, would not be able with certainty to say which was the globe, which the cube." And later, George Berkeley (1707) said similarly that we should expect such a man not to know whether anything was "high or low, erect or inverted . . . for the objects to which he had hitherto used to apply the terms up and down, high and low, were such only as affected or were in some perceived by tough; but the proper objects of vision make a new set of ideas, perfectly distinct and different from the former, and which can in no sort make themselves perceived by touch." Berkeley goes on to say that it would take a long time to associate the two. But, contrary to the expectations of the philosophers, we found that directly after the first operation, S. B. could see things immediately that he knew from his earlier touch experience; although for many months, and indeed years, he remained effectively blind for things he had not been able to explore by touch. So Berkeley's assumption that vision and touch are essentially separate is not correct; knowledge based on touch is very important for vision. Most dramatically, S. B. could immediately tell the time by sight from a wall clock on the hospital ward; as he had read time by touch from the hands of his pocket watch, from which the glass had been removed so that he could feel its hands. Even more surprising: following the operation he could immediately read uppercase, though not lowercase letters. It turned out that he had learned uppercase, though not lowercase, letters by touch as a boy at the Blind School from uppercase letters engraved on wooden blocks. The blind children were given only uppercase letters, as lowercase was not used at that time for street signs or brass plates, which it would be useful to read by touching. So the blind school had inadvertently provided the needed controlled experiment, which suggested that active exploration is vitally important for the development of meaningful seeing in children.

Most moving, and most informative, was S. B.'s response to seeing a lathe (which he knew from descriptions) for the first time. Shortly after leaving the hospital, we showed him simple lathe in a closed glass case at the science museum. Though excited by interest, he made nothing of it. Then, with the cooperation of the Museum staff, we opened the case to let S. B. touch the lathe. As reported at the time (Gregory, 1974):

We led him to the glass case, which was closed, and asked him to tell us what was in it. He was quite unable to say anything about it, except that he thought the nearest part was a handle. (He pointed to the handle of the transverse feed.) He complained that he could not see the cutting edge, or the metal being worked, or anything else about it, and appeared rather agitated. We then asked a Museum Attendant for the case to be opened, and S. B. was allowed to touch the lathe. The result was startling; he ran his hands deftly over the machine, touching first the transverse feed handle and confidently naming it as a "handle," and then on to the saddle, the bed and the head-stock of the lathe. He ran his hands eagerly over the lathe, with his eyes shut. Then he stood back a little and opened his eyes and said: "Now I've felt it, I can see."

S. B.'s effective blindness to objects he did not know as remarkably similar to clinical agnosia, and to Ludwig Wittgenstein's (1953) notion of "Aspect Blindness." In our own experience (or rather lack or it) of ambiguous figures, such as Jastrow's Duck-Rabbit—while it is accepted as a rabbit, the duck features are scarcely seen, disappearing into aspect blindness. This is also dramatic in Rubin's Face-Vases, which disappear in turn, sinking into the ground of the invisibility of aspect blindness, to emerge from nothing as materializing figures. Thus Wittgenstein (1953, p. 213) asks of an imaginary aspect-blind person, presented with the reversing-skeleton Necker Cube figure:

> Ought he to be unable to see the schematic cube as a cube? For him it would not jump from one aspect to another. The aspect-blind will have altogether different relationship to pictures from ours.

We found that S. B. did not experience reversals of these (to us) ambiguous figures. For him they were meaningless patterns of lines, and, in general, pictures were hardly seen as representing objects. From this, I suggest (Gregory, 1981) that perceptual phenomenon of ambiguity should be highly useful for investigating meaning and understanding.

There was evidence that he learned to conceive and perceive space, not only by handling objects but also by walking. In the hospital ward he was able to judge distances of objects such as chairs with remarkable accuracy. But looking down from the window—which was some 40 or more feet high—he described the distance of the ground as about his own body height. He said that if he hung from the windowsill with his fingers, he feet would just touch the ground. Blind people avoid jumping down for they do not know what is (if anything!) below them; they feel carefully with their feet first. So he would have had little or no experience of distances below his feet, except for stairs and occasionally ladders. We may conclude that experience of walking was necessary for seeing distance. This is borne out by our, normal, loss of Size Scaling looking down from a high building, when cars and people and so on look like toys, though for the same horizontal distance they look almost their "correct" sizes.

All this is evidence that perception depends neurally on reading or interpreting sensory signals in terms of experience and knowledge, or by assumptions (which may, however, be wrong and misleading to produce illusions (Gregory, 1968, 1980)) of the object world. The Exploratory aim is to amplify and extend first-hand experience to enrich perception and understanding for children and throughout adult life. The effectiveness of the hands-on approach for teaching has been questioned. But in any case, surely capturing interest is the first essential for more formal methods to be effective. It is hard to believe that learning has to be serious; it is far more likely that play is

vitally important for primates to learn how to exist in the world in which they find themselves. It is fascinating to watch children and adults in this play-experiment situation of individual discovery. Although research is needed to be sure, they certainly give every indication of thinking and learning by doing.

It seems that children do not approach questions or experiments from a vacuum; they generally have performed ideas, which may not be appropriate or coherent, but may be held robustly. They may be discovered (both by their parent or teacher) by setting up predictions. Thus in the Exploratory, experiments with gyroscopes, or the Bernoulli effect, are highly surprising and so reveal erroneous conceptions. Assumptions may of course also be discovered through questioning, and spontaneous questions may reveal how children or adults see, or think they see. According to Jean Piaget and several other authorities, young children hold magical notions of cause, not distinguishing between their own responses and the behavior of inanimate objects, and they tend to hold Aristotelean notions of physics of motion and forces. In 1929, Piaget described children as believing that all objects capable of movement—such as bicycles, and the sun and moon—are alive. And Piaget reported many investigations on perception of conservation (or lack of conservation) of matter, finding that most children before the age of 9, when given various shapes of a lump of clay, do not appreciate conservation of substance. Presumably hands-on experience tends to correct such errors; but how good are adults? A marketing trick is to use odd-shaped bottles to make the contents look larger, which fools most people.

Do children, if implicitly, apply the scientific method to generate their understanding of the world? This was the view of Jean Piaget (1896-1980), the greatest name in the field. Piaget came to favour of an outright empiricism, where logic itself is learned. In *The Child and Reality* (1972), Piaget proposes the following hypothesis (p. 94):

> (a) That at every level (including perception and learning), the acquisition of knowledge supposes the beginning of the subject's (child's) activities in forms which, at various degrees, prepare logical structures; and (b) therefore that the logical structures already are due to the coordination of the actions themselves and hence are outlined the moment the functioning of the elementary instruments are used to form knowledge.

Piaget offers experiments to show effects of inferences during perceptual development in children, showing that perceptions change as inferences change. For example (*The Child and Reality*, p. 95): "A young child is shown briefly two parallel rows of four coins, one being spaced out more than the other: The subject will then have the impression that the longer row has the more coins." Piaget goes on to say that joining the corresponding coins of each row by lines, or joining them in other ways, has different effects for different ages or stages of perceptual development. So Piaget suggests that different inferences about the lines are made, each making the rows of coins appear somewhat different. He also cites an experiment from his laboratory in which the numbers 1 and 7 are shown with their tops hidden, and at different orientations. When the 1 is tilted to the slope of the 7, it is still read as a 1 when ending a sequence likely to be a 1, but otherwise it is seen as a 7. So probability affects perception in children.

Older children's notions are reported in *Children's Ideas in Science*, edited by Rosalind Driver, Edith Guesne, and Andree Tiberghien (1985). This starts with an account by Rosalind Driver of two 11-year-old boys in a practical class measuring the length of a suspended spring, as equal

weights are added to a scale pan. In the middle of the experiment one of the boys unlocked the clamp and moved the top of the spring up the retort stand. He explains:

> This is farther up and gravity is pulling it down harder the farther away. The higher it gets the more effect gravity will have on it because if you just stood over there and someone dropped a pebble on him, it would just sting him, it wouldn't hurt him. But if I dropped it from an airplane it would be accelerating faster and faster and when it hit someone on the head it would kill him.

This reveals the boy's view of gravity, which is not quite ours.

Whether young children ask abstract or philosophical questions has been asked by an American teacher of philosophy, Gareth Matthews in *Philosophy and the Young Child* (1980). As an example, a boy who had often seen airplanes take off, disappearing in the distance, flew for the first time at the age of 4 years. After takeoff, he turned to his father and said in a puzzled voice: "Things don't really get smaller up here."

How do children come to derive reality from appearances? Is a single dramatic experience such as flying for the first time—or discovering that patterns of spectral lines from glowing gases correspond to light from the stars—sufficient for a paradigm change of view or understanding in children? Can adults go back to the drawing board to see the world afresh?

For looking at the details of how perception works, it is convenient to consider somewhat separately the early stages of how patterns and colors are signaled by the retina and analyzed by the initial stages of the brain's perceptual systems, and then the cognitive (knowledge-based) processes of selecting and testing perceptual hypotheses of the objects and situations that we have to deal with to survive. A particular question that concerns us—and we have no clear answer—is how the various signaled features finally come together, without obvious discrepancies. For example, given that color and brightness are signaled by different parallel systems, why don't they lose their registration to separate and produce spurious edges at borders of objects?

Curiously, our mammalian ancestors did not have effective color vision before the primates, including ourselves at the top of the evolutionary tree. So it might be expected that for us brightness contrast is more significant than color contrast for recognizing objects, and this is generally so. The importance of brightness rather than color contrast is clear from the effectiveness of black and white photography. Switching out the color of a TV set does little to impair our perception (apart from watching snooker) except in rather special, though sometimes biologically important, situations. From this simple experiment we can see that color is useful for spotting red berries in green foliage, seeing through camouflage, remotely sensing the edibility of fruit and meat, which could be a major reason why color vision developed in primate evolution. It had already developed, in various forms, in insects, fishes, and birds, but curiously it was lost for mammals, to be reinvested in our immediate primate ancestors.

In some of our experiments, we do the converse of switching out the color of a TV set: we remove brightness differences while preserving color contrast. This gives "isoluminant" displays, which can be seen only by color vision because there are no brightness differences. We have developed several techniques for producing color-without-brightness contrast, usually for a pair of colors, such as red and green. It is important to ensure that they are set to equal brightness for each observer, for there are individual differences of color sensitivities which, when extreme, are

color-blindness (or better, "color anomaly") which is usually reduced sensitivity to (so-called) red or green light. For these experiments it is important that neighboring color regions do not overlap, or have gaps, because such registration errors would produce brightness differences at the color borders. So producing truly isoluminant displays presents some technical problems (and it rarely occurs in nature), but some of the phenomena can be seen in formal color printing when the print has the same brightness as its different-color background. When the print and background have the same brightness, it is difficult to read and the edges of the letters appear "jazzy." The print is unstable, moving around disconcertingly. In spite of the loss of stability, and uncertainty of just where the edges are, there is hardly any loss of visual acuity as measured with a grating test, although letters are more difficult to read. The fact that letter acuity though not grating acuity is impaired suggests that precise *position* of edges (called "phase" information) is lost at isoluminance, though *separations* between nearby features are signaled almost normally. Reading is particularly difficult when letters are closely spaced. They can also lose their individual identities, breaking up into unfamiliar units.

Losses may also be of neurally higher-level brain processes. Most striking is the appearance (or rather, disappearance) of an isoluminant face. This can be shown best with a matrix of red and green dots as in coarse screen printing: when the two colors are set to isoluminance, the face immediately loses all expression and looks flat, with meaningless holes where the eyes and mouth should be. It no longer looks like a face: it becomes meaningless shapes. Although this is a "subjective" observation, it is unmistakable. It is very strong evidence of drastic perceptual loss when only color is available, for almost anything is normally accepted as a face. This, indeed, makes the cartoonist's work possible because just a few lines can evoke an expressive face; so it is remarkable that face perception is so completely lost with isoluminant color contrast. It is important to note that this loss does not occur when a normal brightness-contrast picture is blurred, for example by being projected out of focus, so this loss of face seems to be a central perceptual phenomenon.

The kinds of losses that occur with normal observers at isoluminance are strikingly like the clinical symptoms of amblyopia, or a lazy eye. This "artificial amblyopia" of isoluminance is convenient for experiments because it can be switched on and off and compared with the normal vision in the same individual. Also, we can see what happens and compare our experience with the reports of people who suffer from amblyopia, which is a help for at least intuitive understanding.

A further and dramatic loss is of a certain kind of stereoscopic depth. The American psychologist Bela Julesz discovered, over 20 years ago, that when slightly different random dot patterns are presented, one to each eye, in a stereoscope, regions of dots which are shifted sideways for one eye are seen as lying at a different distance from the rest of the dots which are not displaced. This shows that the brain can compare meaningless dot patterns presented to the eyes and compute depth from small horizontal shifts—which normally occurs for different distances, as the eyes receive slightly different views as they are horizontally separated by a few centimeters. But when the dots are, for example, green on a red background of the same brightness, this stereoscopic depth is lost. We are now comparing this dramatic loss of stereoscopic depth for meaningless dot patterns (which, however, is perhaps never quite complete) with what happens when there are lines and meaningful objects presented in stereoscopic depth to the two eyes. There is some evidence that edges activate different neural mechanisms from the random dots, because a few people have "line" but not "random dot" stereo vision. Perhaps also the meaning, or object-significance, of what is presented may be important in how the brain compares features for perceiving depth.

There is a corresponding phenomenon for movement. When a pair of such random dot figures are alternated, about 10 times/sec, and viewed with one or both eyes, the shifted dot region separates from the rest of the dots and moves right and left. We find that when the dots are set to isoluminance, the displaced dots are lost among the others and no movement is seen (Ramachandran and Gregory, 1978). This is remarkable, because the dots can be quite large, and clearly visible individually, and yet this kind of stereo depth and movement are lost without brightness information.

Visual channels may be isolated in various ways, including selective adaptation to colors (giving colored afterimages); to prolonged viewing of tilted lines (making vertical lines look tilted in the opposite direction); to movement (as in the "movement aftereffect," which was known to Aristotle). We have recently found that continuous *real* movement is signaled by the same neural channel as discontinuous *apparent* (or phi) movement, which may be seen when stationary lights are switched on and off in sequence—provided the gaps in space and time of the apparent movement are not too great (Gregory and Harris, 1984). When the gaps are large (greater than about 10 min arc subtended angle), movement can still be seen, but now it is signaled by a different neural channel, or cortical analyzing system. This we have found by showing that real movement can cancel opposite-direction apparent movement. This is done by illuminating a readily rotating sector disk with stroboscopic short flashes of light set to make it appear to rotate backwards from its true motion, and also with a variable-intensity continuous light. This produces, say, real clockwise movement and, at the same time, apparent anticlockwise movement of the disc. These movements can be set to cancel, or null, but adjusting the relative intensities of the strobe and continuous lights. At the null point there is only a random jitter, with no systematic movement. The null point is not affected by the disturbing effect of adapting to prolonged viewing of movement. The movement aftereffect affects the real and apparent movement equally, which is strong evidence that they are sharing a common channel. The nulling of real against short-range, apparent movement occurs even though the strobe and the continuous lights have different colors, so the eye's three color channels share a common movement system.

There is, however, an interesting limit to the real/apparent-movement shared channel. When the strobe's flash rate is set to give large jumps of the rotating sectors, nulling no longer occurs. The two movements are now seen passing through each other, simultaneously. These observations indicate a shared channel for real- and short-range apparent movement, but a separate channel for long-range movement. It is well known to cartoon film animators that the long-range movement of large jumps between frames has cognitive characteristics, such as being affected by which features are parts of the same object, or are likely to move separately.

An intriguing question is how the various sources of information from different parallel neural channels combine to give unified perceptions of objects. Although neural channels have different characteristics, and in spite of selective adaptations (which affect some channels but not others), and in spite of distortions (which may be dramatic), we do not experience spurious multiple edges. This surely requires some explanation. We suggest that misregistrations are avoided by a process of "border-locking," such that luminance borders pull nearby color edges to meet them (Gregory and Heard, 1979). So spatial registration discrepancies are prevented, although at the cost of some distortions, which may be very evident. Presumably, some visual distortion of size and curvature is not important in nature, although multiple edges, where there should be but one, would be seriously confusing. So, we suggest, registration is maintained by border-locking (where color is slave to luminance) at the cost of some distortion.

It turns out that the classical perspective distortion illusions (such as the Muller-Lyer and the Poggendorf illusions) remain essentially unchanged when presented with their lines having color contrast to their backgrounds, and set to isoluminance (Gregory, 1976). But some illusions, notably the Cafe Wall illusion (Gregory and Heard, 1979), which has no perspective-depth features, appear undistorted when isoluminant. It seems that early sensory processing is affected by isoluminance (as in the parallel lines of the Cafe Wall illusion), but the cognitive reading (or misreading) of perspective depth from converging lines, which can give spatial distortions (Gregory, 1974), is unaffected by isoluminance—it does not matter how the information arrives for cognition.

Recently, David Hubel and Margaret Livingstone (1987) have found strong evidence for separate cortical systems for representing and analyzing luminance and color information. It now seems that color is primarily analyzed by blobs in the third layer of the striate cortex, while orientations, etc., signaled by luminance differences are analyzed by interblob cells at this early stage of visual processing. On a matter of detail, we disagree with one of Hubel and Livingstone's observations, for, as mentioned above, we find that the perspective depth distortion illusions remain at isoluminance; bu they claim that these and all perspective depth disappear. This is not our experience, but no doubt this discrepancy will soon be resolved.

# REFERENCES

Berkeley, G. (1707), A New Theory of Vision (London, 1910).

Driver, R., Guesne, R. D., and Tiberghien, A. (1985), Children's Ideas in Science (Milton Keynes, Open University Press).

Gregory, R. L. (1968), Perceptual Illusions and Brain Models, Proc. Roy. Soc., B, 171, 278-96.

Gregory, R. L. (1976), Visions with Isoluminant Color Contrast: 1. A Projection Technique and Observations. Perception, 6 (1), 113-119.

Gregory, R. L. (1980), Perceptions as Hypotheses, Phil. Trans. Roy. Soc., B, 290, 181-97.

Gregory, R. L. (1981), Mind in Science (London, Weidenfeld & Nicolson).

Gregory, R. L. (1986), Hands-On Science: An Introduction to the Bristol Exploratory (London, Duckworth).

Gregory, R. L. and Harris, J. P. (1984), Real and Apparent Movement Nulled. Nature, 307, 729-730.

Gregory, R. L. and Heard, P. F. (1979), Border Locking and the Cafe Wall Illusion. Perception, 8, 4.

Gregory, R. L. and Wallace, Jean G., Recovery from Early Blindness: A Case Study, Exp. Psychol. Soc. Mongr. II (Cambridge, Heffers). Reprinted in Gregory, R. L. (1974) Concepts and Mechanisms of Perception (London, Meuthen). Pp. 65-129.

Iggo, A. (1982), Cutaneous Sensory Mechanisms, in The Senses, Barlow, H. B. and Mollon, J. D. (eds.), Cambridge, CUP. Pp. 369-406.

Livingston, M. S. and Hubel, D. H. (1987), Psychophysical Evidence for Separate Channels for the Perception of Form, Color, Movement and Depth. J. Neuroscience, 7 (11), 3416-3468.

Locke, J. (1690), Essay Concerning Human Understanding (Ed. P. H. Niddich, 1975, Oxford, OUP).

Matthews, G. B. (1980), Philosophy and the Young Child (Cambridge, MA HUP).

Piaget, J. (1972), The Child and Reality: Problems of Genetic Psychology. Transl. by Arnold Rosin. (London, Frederick Muller.)

Ramachandran, V. S. and Gregory, R. L. (1978), Does Color Provide an Input to Human Motion Perception? Nature, 275, 55-56.

Wittgenstein, L. (1953), Philosophical Investigations, Transl. by G. E. M. Anscombe. (Oxford, Blackwell.)

# THE PERCEPTION OF THREE-DIMENSIONALITY ACROSS CONTINUOUS SURFACES*

Kent A. Stevens
Department of Computer Science
University of Oregon
Eugene, Oregon

## ABSTRACT

The apparent three-dimensionality of a viewed surface presumably corresponds to several internal perceptual quantities, such as surface curvature, local surface orientation, and depth. These quantities are mathematically related for points within the silhouette bounds of a smooth, continuous surface. For instance, surface curvature is related to the rate of change of local surface orientation, and surface orientation is related to the local gradient of distance. It is not clear to what extent these 3D quantities are determined directly from image information rather than indirectly from mathematically related forms, by differentiation or by integration within boundary constraints. An open empirical question, for example, is to what extent surface curvature is perceived directly, and to what extent it is quantitative rather than qualitative. In addition to surface orientation and curvature, one derives an impression of depth, i.e., variations in apparent egocentric distance. A static orthographic image is essentially devoid of depth information, and any quantitative depth impression must be inferred from surface orientation and other sources. Such conversion of orientation to depth does appear to occur, and even to prevail over stereoscopic depth information under some circumstances.

## INTRODUCTION

One can derive a compelling impression of three-dimensionality from even static, monocular surface displays. Figure 1, for example, suggests an undulating surface. The three-dimensionality of this figure can be dramatically enhanced when one removes the visual evidence about the surface on which the figure is printed. If, say, the pattern is viewed on a graphics display, in a darkened room, monocularly and without head movements, the apparent three-dimensionality is particularly vivid, sufficiently so that one could replicate the apparent surface by curving a ruled sheet of paper and holding it in a particular attitude.

On reflection, it is actually quite curious that a pattern of lines such as those in figure 1 provides so fixed and stable a percept. There is, after all, an infinity of possible 3D surfaces containing lines that would project to that 2D pattern. To posit that the pattern corresponds to a *particular* surface requires certain, specific, strongly constraining assumptions. A theory has been developed of the geometric constraints that support such inferential 3D percepts, one that explains how a range of 3D qualities, such as local surface orientation and curvature might be derived in principle (Stevens 1981a, 1983b, 1986). But it is difficult to extend such theories to explain more precisely *what* 3D information is extracted and internally represented in the process of deriving apparent three-dimensionality from such a 2D stimulus. It is one thing to discuss perception in terms of

---

"affordances," "cues," or other characterizations of incident information, and quite another thing to determine the specific course of processing that takes incident information into explicitly represented perceptual quantities.

The remarkable ability to derive surface information from simple monocular configurations has been quite difficult to explain adequately within any of the traditional psychological paradigms. The difficulty stems, I believe, from the lack of basic understanding about what constitutes "apparent three-dimensionality." Depth perception is an often-used term that refers to the perception of surfaces and points in 3D. What differentiates the perception of mere 2D patterns of stimulation from 3D arrangements, seemingly, is perception of the third dimension, namely depth or distance from the viewer to points in space. Gibson insightfully proposed that "visual space perception is reducible to the perception of visual surfaces, and that distance, depth, and orientation...may be derived from the properties of surfaces" (Gibson 1950). To Gibson, the term "apparent three-dimensionality" refers to the perception of more than merely the "third dimension." Visual perception clearly developed to operate in the richly redundant visual world. But the very little 3D information in figure 1 hardly compares to the redundant and seemingly unambiguous wealth of incident information afforded by a natural scene. It might justifiably be relegated to the domain of so-called "picture perception."

Approaches toward understanding surface perception that attempt to isolate the contribution provided by a particular cue, such as texture or contours, or motion or stereopsis, have often been criticized as failing to address enough of the problem. By not embracing the complexity of natural scenes, it is argued, one fails to examine the system in the environment for which it was designed. But while one might well fail to observe important phenomena when only examining components in isolation or in simple combination, by *not* doing so one might equally fail to observe effects central to the strategies that allow the system to effectively deal with complexity and redundancy.

If vision is regarded computationally as the construction of internal descriptions of the visual world, there is no particularly compelling reason to expect qualitatively different modes of visual processing depending on whether the retinal image derives from a picture or a real scene. If one does not expect a different mode for "picture perception," one must then explain how an ambiguous and obviously underspecified 2D stimulus can result in a definite and stable 3D percept.

The challenge, then, is to understand our seeming ability to perceive more specifically than is objectively specified by the stimulus. To Helmholtz, Gregory, and others, this ability stems from the basic perceptual strategy of "unconscious inference." To mix terminology from traditionally antagonistic schools of thought on this matter: higher-order variables in the incident optical array are cues that afford particular 3D inferences. After a while such word play is seen for what it is, and we should go on to more constructive explorations. Substantial progress will likely come only with understanding of the nature of the 3D percept, something that has been given remarkably little attention over the entire history of perceptual studies.

As will be discussed, this task is difficult in theory, because of various mathematical equivalences among different representational forms, and difficult in practice, because of the robustness of the visual observer in performing psychophysical judgments. Despite the intrinsic difficulty, however, there is some evidence that surface perception is sufficiently modular and restricted in its ability to extract and combine 3D information as to be amenable to study using traditional psychophysical methods.

# QUANTIFYING APPARENT THREE-DIMENSIONALITY

Following the usage by Foley (1980), *absolute distance* will refer to the egocentric range from an observer to a specific 3D point, which might be a point on a visible surface. *Relative distance* refers to a ratio of absolute distances (without knowing the absolute distances, one might know that one distance is twice another). In this usage *depth* refers specifically to the *difference* of absolute distances to a given point and a reference point. (Hence the depth of a given point relative to a reference point might be known in absolute units without knowing the overall absolute distances involved. Also, if the depth at a point were known and the absolute distance to the reference point were known, their algebraic sum would specify the absolute distance to the given point.)

In addition to scalar distance information at a point, derivatives of distance information specify the orientation of the tangent plane and about curvature of the surface in the vicinity of a point. Surface orientation has two degrees of freedom, and is readily described as a vector quantity related to the normal to the tangent plane (Stevens 1983c). The psychological literature has long used the magnitude quantity *slant* to refer to the angle between the line of sight and the local surface normal (slant varies from 0 to 90°). The other degree of freedom, the *tilt* of the surface, specifies the direction of slant, which is the direction to which the normal projects onto the image plane, and also the direction of the gradient of distance (Stevens 1983a). Since the slant-tilt form aligns with the direction and magnitude of the local depth gradient, it provides many advantages for encoding surface orientation, such as allowing for simultaneous representation of precise tilt and imprecise slant, being closely related to various monocular cues such as shading, texture foreshortening, motion parallax, and perspectivity, and providing for (Necker-type) ambiguity in local surface orientation as reversals in tilt direction (see Stevens, 1983c).

Derivatives of surface orientation, or higher derivatives of distance, are related to surface curvature (across a continuous, twice-differentiable region). Surface curvature also has two degrees of freedom in the neighborhood of a surface point, which might be encoded as principle curvatures, or their image projections.

The central problem, which I will illustrate momentarily, is that across a continuous surface it is possible to convert among these different forms by differentiation (in one direction) and integration (in the other). One source of information about local slant might be used to infer both surface curvature and depth, and another might indicate curvature information directly. With sufficient boundary constraints the information provided by any source might be converted to a form comparable with another across a continuous surface. In general, then, it is difficult to determine whether a given 3D quantity M is derived directly from the image or indirectly from derivatives or integrals of M.

The mathematical equivalences among these various forms of 3D information leave quite open the empirical question of to what extent surface curvature is registered directly versus converted internally (Stevens 1981b; Cutting and Millard 1984; Stevens 1984), and furthermore, the question of the extent to which this information is represented quantitatively rather than qualitatively (Stevens 1981a, 1983b, 1986).

# THE 3D INFORMATION CONTENT OF A SIMPLE STIMULUS

Returning to figure 1, what sorts of 3D information can be extracted feasibly? Observe that it consists merely of a family of parallel curves, interpreted as the orthographic projection of parallel curves across a continuous surface. Given the nature of orthographic projection, this pattern is devoid of information about the third dimension (distance). And yet, one sees measurable depth as well as slant in monocular stimuli consisting of line-drawing renditions of continuous ruled (developable) surfaces (Stevens and Brookes, 1984a). Both orthographic (as in figure 1) and perspective projection were used. Using a randomized-staircase forced-choice paradigm, apparent slant was measured by varying the aspect ratio of an ellipse that was briefly superimposed on the monocular surface stimulus. Observers readily interpreted the ellipse as a foreshortened circle slanted in depth, and by adjusting the aspect ratio it could be made to appear flush on the surface. The resulting slant judgments were in close correspondence to the predicted geometric slant of the stimuli.

The apparent depth in these stimuli was then tested by superimposing a stereo depth probe over the monocular surface. Apparent depth was probed stereoscopically using a device similar to Gregory's (1968, 1970) "Pandora's Box." A Wheatstone-style stereoscope provided near-field (38 cm) convergence and accommodation, well within the range of acute stereopsis. After first fixating a binocular point on an empty field, the monocular stimulus was presented briefly (for as little as 100 msec) to the dominant eye only, after which a binocular probe was superimposed at a given stereo disparity over the monocular stimulus for an additional brief interval. Subjects performed a randomized-staircase forced-choice experiment in which the depth of the stereo probe was compared with that of the monocular surface at various locations. Just as Gregory (1970) found measurable apparent depth in a variety of illusion figures, minimal renditions of monocular surfaces, such as figure 1, are also perceived quite measurably in the third dimension.

The experiments suggest that in orthographic projection the visual system can compute from local surface orientation a depth quantity that is commensurate with the relative depth derived from stereo disparity. Apparent slant is a measure of the local gradient of depth, i.e., the rate of change of depth (and being the derivative of distance, slant is independent of the absolute distance to the surface). Depth might be integrated from slant across the surface, but only up to a constant of integration. How, then, are monocular and stereo depth coupled so that they can be compared? The perceptual assumption used to link these two spaces, apparently, is that the absolute distance of the monocular surface at the given fixation point equals that of the stereoscopic horopter at that point. This hypothesis seems sound in that whatever surface location is fixated in sharp focus is likely to lie at zero disparity, since in the near field at least, there is close coupling between vergence and accommodation that brings into sharp focus the (zero disparity) fixation point. The fixated point (seen monocularly in our stimuli but binocularly in normal vision) is thus assumed to be at the absolute distance of the horopter. With the two depth measures sharing a common zero intercept, monocular depth from slant, appropriately scaled by the reference distance, could then be compared to depth from stereo disparity. This conjecture remains to be confirmed empirically.

# DEPTH FROM GRADIENT, CURVATURE, AND DISCONTINUITY INFORMATION

In addition to demonstrating the perception of three-dimensionality from highly underspecified stimuli, these observations suggest to us that the visual system has a robust ability to internally convert one form of 3D information into another mathematically equivalent form. The perception of depth from the various so-called monocular "depth cues" (such as shading, contours, and texture gradients) may well provide "direct" information about surface curvature and shape, and only indirect information about depth.

More generally, we propose that shape properties associated with derivatives of distance, specifically surface orientation, curvature, and loci of discontinuity, both in depth (edge boundaries) and tangent plane (creases), are the primary percepts, and that smoothly varying depth across continuous regions is recovered subsequently and indirectly (Stevens and Brookes, 1987b,c).

This proposal explains various phenomena concerning apparent depth from stereopsis. The apparent depth of an isolated bar or point is predicted quite well by the geometry of the binocular system, with depth a straightforward function of stereo disparity and a reference binocular convergence signal (Foley, 1980). But various depth phenomena have been reported recently in the perception of more complicated surface-like stimuli that are not predicted by such a direct functional relationship (Gilliam et al., 1984; Mitchison and Westheimer, 1984). Gilliam et al. (1984) argue that depth derives most readily from disparity discontinuities, and Mitchison and Westheimer (1984) show that coplanar arrangements of lines result in elevated thresholds for depth detection. In a series of experiments in which binocular stimuli presented contradictory monocular and stereo information, we found instances where the stereo information was dramatically ineffective in influencing the 3D percept (Stevens and Brookes, 1987c). The patterns were line-drawn stereo depictions of planar surfaces, rendered orthographically and in perspective, and devoid of disparity discontinuities and disparity contrast (e.g., with a surrounding frame or background). Constant gradients of stereo disparity, consistent with slanted planes, were introduced that were orthogonal to or opposite to the monocularly suggested depth gradients. The monocular interpretation dominated in judgments of apparent surface slant and tilt and in 2-point relative depth ordering. Figure 2, for example, is a stereogram of coplanar lines, with disparities varying linearly in accordance with a slanted plane. The dominant depth impression is the monocular interpretation of a perspective view of a corridor extended in depth.

We hypothesize that stereo disparity influences the monocular 3D interpretation primarily where the distribution of disparities indicates surface curvature and depth discontinuities (i.e., where disparity varies discontinuously or has nonzero second spatial derivatives). Stereo depth across surfaces is substantially a reconstruction from disparity contrast, analogous to brightness from luminance contrast. Consistent with this conclusion are a variety of depth-contrast effects in stereopsis, such as a brightness-contrast analogue in depth (Stevens and Brookes, 1987b), a Craik-O'Brien-Cornsweet analog (Anstis et al., 1978), and various depth induction effects (e.g., Werner, 1938).

# REFERENCES

Anstits, S.; Howard, I. P.; and Rogers, B.: A Craik-O'Brien-Cornsweet Illusion for Visual Depth. Vision Research, vol. 18, 1978, pp. 213-217.

Cutting, J. E.; and Millard, R. T.: Three Gradients and the Perception of Flat and Curved Surfaces. Journal of Experimental Psychology: General, vol. 113, 1984, pp. 217-220.

Foley, J. M.: Binocular Distance Perception. Psychological Review, vol. 87, 1980, pp. 411-434.

Gibson, J. J.: The Perception of the Visual World. Boston: Houghton Mifflin, 1950.

Gilliam, B. J.; Flagg, T.; and Finlay, D.: Evidence for disparity change as the primary stimulus for stereoscopic processing. Perception and Psychophysics, vol. 36, 1984, pp. 559-564.

Gregory, R.: Visual Illusions. Scientific American, vol. 218, no. 11, 1968, pp. 66-76.

Gregory, R.: The Intelligent Eye. London: Weidenfeld and Nicolson, 1970.

Mitchison, G. J.; and Westheimer, G.: The Perception of Depth in Simple Figures. Vision Research, vol. 24, 1984, pp. 1063-1073.

Stevens, K. A.: The Visual Interpretation of Surface Contours. Artificial Intelligence, vol. 217, Special Issue on Computer Vision, 1981a, pp. 47-74.

Stevens, K. A.: The Information Content of Texture Gradients. Biological Cybernetics, vol. 42, 1981b, pp. 95-105.

Stevens, K. A.: Surface Tilt (the Direction of Surface Slant): A Neglected Psychophysical Variable. Perception and Psychophysics, vol. 33, 1983a, pp. 241-250.

Stevens, K. A.: The Line of Curvature Constraint and the Interpretation of 3-D Shape from Parallel Surface Contours. Eighth International Joint Conference on Artificial Intelligence, August 1983b.

Stevens, K. A.: Slant-Tilt: The Visual Encoding of Surface Orientation. Biological Cybernetics, vol. 46, 1983c, pp. 183-195.

Stevens, K. A.: On Gradients and Texture "Gradients." Commentary on: Cutting and Millard, 1984. Three Gradients and the Perception of Flat and Curved Surfaces. Journal of Experimental Psychology: General, vol. 113, 1984, pp. 217-220.

Stevens, K. A.: Inferring Shape from Contours Across Surfaces. In: From Pixels to Predicates: Recent Advances in Computational Vision. A. P. Pentland, ed., Norwood, NJ: Ablex, 1986, pp. 93-110.

Stevens, K. A.; and Brookes, A.: Probing Depth in Monocular Images. Biological Cybernetics, vol. 56, 1987a, pp. 355-366.

Stevens, K. A.; and Brookes, A.: Depth Reconstruction in Stereopsis. Proceeding of the First IEEE International Conference on Computer Vision, London, June 1987b.

Stevens, K. A.; and Brookes, A.: Integrating Stereopsis with Monocular Interpretations of Planar Surfaces. Vision Research, vol. 28, no. 3, 1988, pp. 371-386.

Werner, H.: Binocular Depth Contrast and the Conditions of the Binocular Field. American Journal of Psychology, vol. 51, 1938, pp. 489-407.
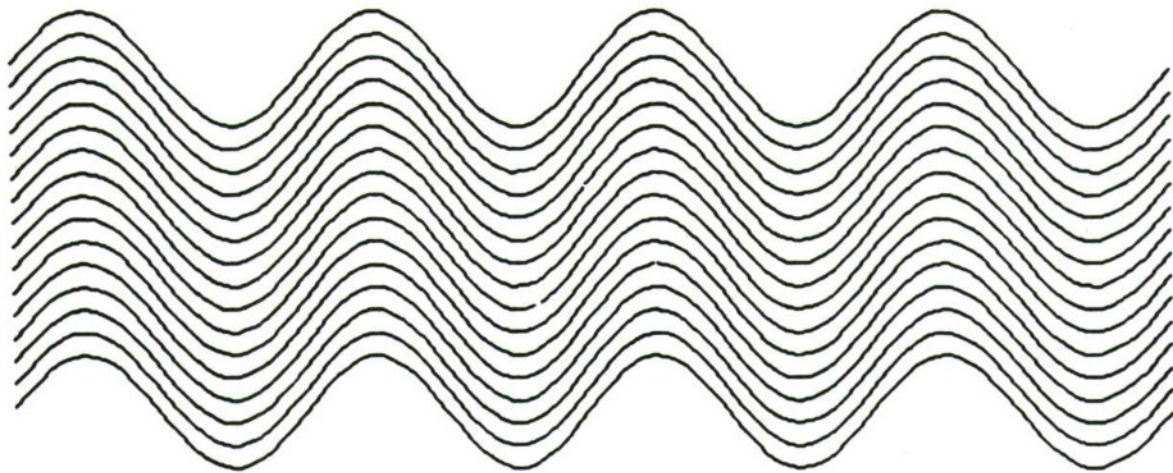
Figure 1.– Undulating lines.
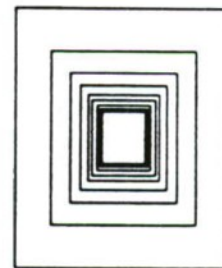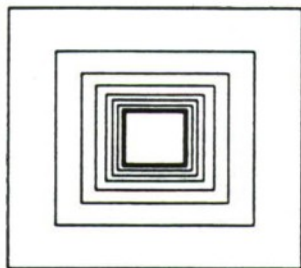


Figure 2.– Stereogram of coplanar lines.

# PERCEIVING ENVIRONMENTAL PROPERTIES FROM MOTION INFORMATION: MINIMAL CONDITIONS

Dennis R. Proffitt
University of Virginia
Charlottesville, Virginia

Mary K. Kaiser
NASA Ames Research Center
Moffett Field, California

## INTRODUCTION

Everyday perception occurs in a context of nested motions. Eyes move within heads, heads move on bodies, and bodies move in surroundings that are filled with objects, many of which can themselves move (Gibson, 1966). Motion is omnipresent in perception. Stabilize an image on the retina and it rapidly becomes imperceptible (Pritchard, 1961). Not only is motion a necessary condition for perception, but it is also a sufficient condition for the perception of a variety of environmental properties.

Until recently, spatial instruments had few degrees of freedom with respect to the sorts of motion-carried information that they could provide. With increasing opportunities to employ animation, spatial instruments can be crafted that are tied less to artificial conventions and more to the natural condition of everyday perceptual experience.

The implications of perception research for display design derive from the methods employed by visual scientists in their investigations of how people extract environmental properties from optical information. The approach taken in perception research involves a seeking of *minimal stimulus conditions* for perceiving these properties. Stimuli that typically evoke relevant perceptions are decomposed into minimal information sources, and these sources are evaluated separately. It is almost always found that we humans rely on a large variety of information sources in perceiving any particular aspect of the environment. Knowledge of minimal conditions for perceiving environmental properties can be utilized in the design of effective and technologically efficient spatial instruments.

Since motion information is a minimally sufficient condition for perceiving numerous environmental properties, its use in spatial instruments eliminates the need to employ most of the conventions typically found in static displays. Moreover, in some contexts animated displays can elicit more accurate perceptions than are possible for static displays.

In this chapter, we discuss the status of motion as a minimal information source for perceiving the environmental properties of surface segregation, three-dimensional (3-D) form, displacement, and dynamics. The selection of these particular properties was motivated by a desire to present research on perceiving properties that span the range of dimensional complexity.

# SURFACE SEGREGATION

Surface segregation refers to the separation of distinct surfaces in depth. In order to represent surface segregation on a two-dimensional (2-D) display surface, the surfaces must be distinguished by some apparent optical differences. These distinctions can be achieved with either static images or animated displays; however, only with motion can surface segregation be specified by a single cue without introducing ambiguous depth-order relations. Moreover, the implicit viewer assumptions needed to interpret moving displays are derived from the laws of dynamics, and thus are more fundamental in nature than are those accessed in interpreting static displays.

## Perceiving Surface Segregation in Static Images

In pictures, surfaces are typically distinguished by color contrasts produced by differences in intensity or wavelength. One surface thereby becomes separated from another at an edge. Figure 1 depicts the familiar faces-vase figure introduced by Rubin (1915). This figure exemplifies the inherent figure-ground ambiguity of all static displays. Here, depending upon which is taken as figure, the vase or the faces, depth-order relations reverse (depth order being a term that refers to what is in front of what).

In order to resolve this depth-order ambiguity, additional cues must be supplied. One effective cue is occlusion. As is shown in figure 2, having one surface appear to be partially covered by another is an effective convention for specifying depth order. It is important to realize, however, that the disambiguation of figure 2 is achieved only through the activation of implicit assumptions or biases on the part of the viewer. The viewer must assume that the apparent far surface does not, in fact, have a notch cut out of it. As the Ames demonstrations on the overlay show, if this assumption is violated, viewers will see erroneous depth-order relations (Ittelson, 1968).

Another static convention that helps to resolve depth-order ambiguity is the use of familiar surfaces. In figure 3, the "A" is typically seen in front of the background surface. As figure 1 showed, what is taken as figure—vases or face–is perceived as being in front of the apparent ground (Rubin, 1915). This perceptual bias can be exploited by representing the intended forward surface with a familiar figure. However, as with occlusion, this convention relies heavily on inherent viewer biases. The A is *assumed* to have been placed atop the surrounding surface, as opposed to having been cut out of it. This assumption may be in error.

The inclusion of additional cues, such as shading, perspective, or solid modeling, will further constrain depth-order interpretations. However, so long as the viewer cannot obtain multiple perspectives on the objects depicted, the display remains inherently ambiguous. Again, the Ames demonstrations serve to show that observers can always be made to have erroneous perceptions whenever they are constrained to view an object from a unique perspective.

Intermediate between static and animated displays are those that include flicker. Wong and Weisstein (1987) found that surface segregation is observed in displays consisting of randomly placed dots when a particular region is made to flicker. Moreover, the flickering region usually appears to be behind adjacent nonflickering regions. Spatial instruments have yet to exploit this perceptual influence of flicker.

## Perceiving Surface Segregation in Motion Displays

The ability of motion information to specify surface segregation without depth-order ambiguity was demonstrated by Gibson et al. (1969). They produced movies of randomly textured surfaces. When the surfaces were superimposed and stationary, segregation could not be achieved. However, when one or both of the surfaces moved, they separated into distinct surfaces and their depth order became unequivocal.

It was thought that the ongoing occlusion of the far surface by the near one served as the essential source of information for the surface segregation demonstration of Gibson et al. Recently, however, Yonas, Craton, and Thompson (1987) showed that surface segregation could be achieved without ongoing occlusion occurring at surface edges. They created a computer-animated display in which surfaces were defined by randomly positioned points of light. As with the original Gibson et al. display, when the simulated surfaces were stationary, there was no information suggesting that more than one surface was present; however, when the surfaces moved, their segregation became apparent. In this case, segregation and depth order were specified by the relative motion of point-lights on different surfaces, and by the disappearance of the lights on the far surface when they passed beneath the subjective contour that defined the edge of the close surface.

There are, of course, implicit assumptions that must be made in interpreting moving displays; however, they are of a fundamentally different sort than those that were discussed for static presentations. For static displays, the assumptions are characterized by notions of likelihood and simplicity. It is highly unlikely that anyone would create a display such as figure 2 with the intent of depicting a square located behind a notched square. Moreover, by any criterion of simplicity, the obvious interpretation of figure 2 is the simpler of the two (or three) depth-order alternatives (see, for example, Leeuwenberg, 1982). For animated displays, the implicit assumptions reflect fundamental laws of dynamics. Surfaces are not destroyed or brought into being when they pass in front of, or go beyond, more distant surfaces. Unlike those accessed when viewing static displays, the assumptions engaged when perceiving animated displays are based upon dynamical laws.

## THREE-DIMENSIONAL FORM

Any 2-D representation of a 3-D object is inherently ambiguous. This is true of both static and moving displays. The virtue of animated displays, however, is that time can substitute for the lost spatial dimension.

Implicit viewer assumptions are required to recover 3-D relations from either static or moving 2-D projections. As was found for perceiving surface segregation, those engaged when viewing animated displays are grounded in the laws of dynamics as opposed to the conventions of artifice.

# Perceiving 3-D Form in Static Displays

Effective means for representing 3-D objects and scenes were discovered by pictorial artists and evolved over time (Gombrich, 1960). Following Berkeley (1709), these pictorial conventions have come to be called secondary or pictorial depth cues. Researchers are still attempting to discover the invented techniques by which artists produced their compelling spatial effects (Kubovy, 1986).

The list of secondary depth cues is a long one; however, all entries share a common origin in the motivation to overcome the ambiguity inherent in 2-D representations of a 3-D scene. The resolution of ambiguity through the implementation of such conventions as solid modeling, perspective, shading, occlusion, familiarity, and so forth is more apparent than real. Demonstrations, such as those of Ames (Ittleson, 1968), show that perception can always be in error when inferring 3-D structure from a single 2-D projection. The possibility of such errors reflect, in turn, on the processing assumptions made when interpreting static displays. As with surface segregation, assumptions grounded in likelihood and simplicity are prevalent. To these are added various assumptive geometric conventions (Kubovy, 1986).

# Perceiving 3-D Form in Motion Displays

The use of geometry can show that the changing spatial pattern, produced when the image of a rotating rigid object is projected onto a 2-D surface, uniquely defines the 3-D configuration of the object. In addition, three projected images of four non-coplanar points undergoing rotation defines the minimal condition for the recovery of structure from motion (Ullman, 1979).

Wallach and O'Connell (1953) showed that people are able to recover 3-D form when viewing 2-D projections of rotating objects. They constructed wire forms and projected their shadows onto screens. Viewers of these shadows reported that they saw only 2-D configurations of lines when the wire forms were stationary; however, they accurately reported on the 3-D configurations when the forms were continuously rotated. Wallach and O'Connell called their demonstration the Kinetic Depth Effect, or KDE.

Interest in KDE has grown over the years. Braunstein (1962), Doner, Lappin, and Perfetto (1984), Todd (1982), and many others have investigated the psychophysics of the phenomenon. Recently, a good deal of research has been directed toward the rigidity assumption.

Recall that transforming a 2-D projection of a rotating form is unique to the form's 3-D configuration only so long as the form remains rigid. Psychologists are much in doubt as to whether the human perceptual system actually implements a rigidity assumption when extracting structure from motion in KDE (Hochberg, 1986).

When the veracity of interpretive assumptions is evaluated, the issue of whether people utilize a rigidity assumption is less important than that such a dynamical assumption is capable of serving as the sole basis for the recovery of structure from motion. Unlike the assumptions embodied in pictorial depth cues, the rigidity assumption is grounded in the following kinematic law: Objects do not distort when rotated. Our perceptual systems were formed in the context of natural constraints. The exploitation of these constraints does not require that they be embodied. The fundamental assumptive nature of the rigidity principle is not based upon whether or not it has been

internalized by the perceptual system, but rather upon this fact: Vision evolved in a context in which this rigidity assumption is inviolate.

It must be conceded that, in a few known circumstances, the assumptions of picture perception interact with those engaged by motion perception. Ames created a trapezoidal surface that looked like a rectangular window viewed at an angle. When observers viewed it monocularly as it underwent rotation, they typically reported seeing an oscillating rectangular window rather than a rotating trapezoid (Ittelson, 1968). It is important to note that this event's 2-D projection is, in fact, inconsistent with the rectangular percept; however, the strong influence of such pictorial assumptions as likelihood and simplicity outweigh, in this case, the motion-carried information defining the actual configuration.

Perceiving 3-D structure from motion information has also been shown to occur for jointed objects. Johansson (1973) placed point-lights on the joints of people and filmed them as they performed actions in the dark. When shown to observers, these movies were readily perceived as depicting people. It was later found that between 0.1 and 0.2 sec was a sufficient exposure duration for perceiving the human form in these films (Johansson, 1976).

Computational theorists have developed effective algorithms for extracting structure from these jointed events, given certain constraints on the motions of the walkers (Hoffman and Flinchbaugh, 1982; Webb and Aggarwal, 1982). These computational models implement assumptions about the local rigidity of moving limbs. In essence, the models assume that the act of rotating or translating a rod (bones in the case of point-light walkers) does not, itself, change the rod's length. This assumption is based upon a kinematic law of nature. The perceptual system may or may not have internalized this law (Proffitt and Bertenthal, 1988); however, it certainly evolved in a world that is governed by it.

## DISPLACEMENT

The motion of an object relative to an observer is referred to as its displacement. Displacement information can be conveyed in static displays only through the use of very artificial conventions. In moving displays, displacement information is presented directly in the natural medium of time. In addition, the perceptual system effectively segregates those motions specifying form from those that define observer-relative displacement.

### Perceiving Displacement in Static Displays

It is not difficult to represent in a static display the fact that an object is moving. What is difficult to represent is the future position that an object will achieve over time. Static representations of motion properties must rely on highly stylized conventions, the most prominent being vector depictions, such as those shown in figure 4. Interpreting such displays not only requires one to effectively read the intended meaning of the conventions, but he or she must also be able to mentally perform the transformation suggested in the representation. People are not very good at such tasks. In fact, when people attempt to extrapolate the future position of moving objects that become occluded behind barriers, they make sizable errors, particularly for complex motion functions (Jagacinski, Johnson, and Miller, 1983).

## Perceiving Displacement in Motion Displays

It is rare in nature for an object to undergo a pure observer-relative translation such that every object point moves with exactly the same motion. In fact, only when objects move in horizontal circles around the observer do common linear motions project to the observer's point of observation; all nonorthogonal distal translations project a rotational component to the observer's viewpoint. The perceptual system deals effectively with complex motions by analyzing them into relative and common motion components (Johansson, 1950). To illustrate this analysis, consider the perception of a rolling wheel.

As is depicted in figure 4, except for the hub, every point on a rolling wheel follows a complex trajectory belonging to the family of cycloidal curves. These trajectories are referred to as the event's absolute motions. The perceptual system segregates these motions into two components, relative rotations and a common-observer relative displacement (Proffitt, Cutting, and Stier, 1979). This perceptual analysis selects the configural centroid as the center of relative rotations. Thus, for a rolling wheel, rotations are seen as occurring about the wheel's hub, and the common motion is seen as the hub's translation. However, if point-lights are attached to an unseen rolling wheel and the configural centroid of these lights does not correspond to the wheel's hub, then a different common motion is seen. Again, relative motions are seen as rotations about the configural centroid, but the common motion is, in this case, the prolate cycloidal path followed by this abstract centroid. This perceptual analysis has also been found to occur for configurations moving in depth (Proffitt and Cutting, 1979). It has been proposed that the selection of the configural centroid, as the center for perceived relative motions, reflects a perceptual preference to minimize relative motions; in centroid relative rotations, all instantaneous relative motions sum to zero (Cutting and Proffitt, 1982).

Research findings on the perceptual analysis of absolute motions into relative and common components have two implications for display design. First, object configuration interacts with displacement perception. Whenever an object undergoes a complex motion, its configural properties influence the common motions that are observed. Although the effects are somewhat different, robust configural influences have also been shown to occur in stroboscopically presented apparent motions (Proffitt et al., 1988). Second, relative and common motions have different perceptual significances (Proffitt and Cutting, 1980). As is depicted in figure 5, relative rotations are used to perceptually define 3-D form, whereas common motions are residual to form analysis, and define observer relative displacements.

## DYNAMICS

The laws of dynamics place constraints on the sorts of motions that can occur in nature. Given these constraints, the patterns observed in natural motions reflect back upon underlying dynamical properties. The motions of colliding objects are a good example of this reciprocal specification of dynamic and kinematic properties.

When objects collide, the laws of linear momentum conservation state that post-collision motions must preserve the event's pre-collision momentum. (For the sake of simplicity, we

exclude considerations of friction and damping.) Given these laws, it can be shown that the ratio of masses for the objects involved in a collision are specified by ratios in their velocities (Runeson, 1977). It has been found that people are relatively good at judging mass ratios when observing collisions (Todd and Warren, 1982; Kaiser and Proffitt, 1984). In addition, people are able to accurately discriminate possible collisions from those that violate dynamical principles (Kaiser and Proffitt, 1987a).

These results do not necessarily imply that the human perceptual system has internalized physical conservation laws, and in fact, the results of recent studies strongly suggest that such laws are not inherent to perceptual processing (Gilden and Proffitt, 1989). However, as has been previously discussed for surface segregation and form perception, our sensory systems need not embody natural laws in order to take advantage of the fact that they evolved in an environment in which dynamical laws are always upheld. Motion information is fundamental because dynamical constraints shaped the natural environment in which vision evolved.

The interpretation of static displays require processing rules shaped in the context of pictorial conventions. The conceptual heritage of static information-processing rules is reflected in their subservience to cognitive beliefs. People hold inaccurate common-sense views about natural dynamics. These erroneous beliefs are reflected in their judgments of static, but not moving, displays.

## Perceiving Dynamics in Static Displays

Recently, an intriguing literature has developed on people's naive beliefs about the laws of dynamics. Called "intuitive physics" by McCloskey (1983), these beliefs influence people's predictions about natural motions; moreover, they are often at odds with the laws of dynamics.

Figure 6 shows one of the problems used by McCloskey, Caramazza, and Green (1980). Depicted is a C-shaped tube that is lying flat on a horizontal surface. A ball is rolled through the tube, and upon exiting, the ball rolls across the surface. Subjects were asked to predict the path taken when the ball exited the tube. Approximately 45% of the undergraduate subjects who were asked this question incorrectly stated that the ball would continue to follow a curved path. McCloskey and his colleagues have conducted numerous similar experiments, all showing that judgments made about natural object motions often reflect erroneous beliefs.

All of these studies required people to make judgments while looking at pictures. The influence of intuitive physics beliefs is pervasive only in such static contexts. These beliefs have been found to have little or no effect on the perception of animated displays.

## Perceiving Dynamics in Motion Displays

We replicated McCloskey et al.'s finding with the C-shaped tube problem, using a design in which observers were asked to judge which of a set of drawn trajectories appeared correct. Then, using the same design, we showed observers animated simulations of balls rolling through C-shaped tubes. Upon exiting the tubes, the balls followed a variety of paths. We found that people almost always chose as correct the natural trajectory when viewing these moving displays, and judged their erroneous predictions as being anomalous (Kaiser, Proffitt, and Anderson, 1985).

We have demonstrated this superiority of motion displays to evoke accurate dynamical judgments in other contexts (Kaiser and Proffitt, 1987b).

Static representations elicit intuitions that reflect cognitive beliefs. Obviously, people would have great difficulty getting about in the world if their perceptions were always tied to their knowledge of physical principles. A baseball outfielder, for example, would probably never succeed in catching a flyball if he was required to plan his pursuit using only his knowledge of physics.

Everyday perceptions necessarily occur in a context of naturally constrained motions. In such circumstances, our perceptual systems can function without recourse to memorial conceptions. Perception is good in motion context because motion is fundamental to the rules of perceptual processing.

## CONCLUSIONS

Motion is an effective source of information for perceiving a variety of environmental properties. Because it is a minimally sufficient information source, it need not be simply added to the conventions employed in static displays. Rather, motion can replace many of these conventions, and in some contexts, motion can elicit more accurate perceptions than are possible for static displays.

Motion information is fundamental to everyday perception. The interpretive assumptions required to extract structure from motion are based upon the laws of nature—i.e., natural dynamics—whereas those evoked by static displays are based upon the artificial conventions of pictorial representations. The advantage that motion displays have over static ones derives from the heritages of the perceptual processes needed for their interpretation. The perceptual processes required to extract structure from motion information were formed in the context of dynamical constraints. The interpretation of static information relies more on perceptual processes that arise with conceptual development, and thus are grounded in such experientially based notions as simplicity, familiarity, and geometrical conventions.

# REFERENCES

Berkeley, G. (1709). An essay towards a new theory of vision. London: Dent.

Braunstein, M. L. (1962). The perception of depth through motion. Psychol. Bull., 59, 422-433.

Cutting, J. E.; and Proffitt, D. R. (1982). The minimum principle and the perception of absolute, common, and relative motions. Cognitive Psychol., 14, 211-246.

Doner, J.; Lappin, J. S.; and Perfetto, G. (1984). Detection of three-dimensional structure in moving optical patterns. J. Exp. Psychol.: Human Percept. Perform., 10, 1-11.

Gibson, J. J. (1966). The senses considered as perceptual systems. Boston: Houghton Mifflin.

Gibson, J. J.; Kaplan, G. A.; Reynolds, H. N.; and Wheeler, K. (1969). The change from visible to invisible: A study of optical transitions. Perception & Psychophysics, 5, 113-116.

Gilden, D. L.; and Proffitt, D. R. (in press). Understanding collision dynamics. J. Exp. Psychol.: Human Percept. Perform.

Gombrich, E. H. (1960). Art and illusion: A study in the psychology of pictorial representation. Princeton, NJ: Princeton.

Hochberg, J. (1986). Representation of motion and space in video and cinematic displays. In K. R. Boff, L. Kaufman, and J. P. Thomas (Eds.), Handbook of perception and human performance. Vol. 1, New York: Wiley, 22:1-63.

Hoffman, D. D.; and Flinchbaugh, B. E. (1982). The interpretation of biological motion. Biol. Cybernetics, 42, 195-204.

Ittelson, W. H. (1968). The Ames demonstrations in perception. New York: Hafner.

Jagacinski, R. J.; Johnson, W. W.; and Miller, R. A. (1983). Quantifying the cognitive trajectories of extrapolated movements. J. Exp. Psychol.: Human Percept. Perform., 9, 43-57.

Johansson, G. (1950). Configuration in event perception, Uppsala, Sweden: Almqvist & Wiksell.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. Perception & Psychophysics, 14, 201-211.

Johansson, G. (1976). Spatio-temporal differentiation and integration in visual motion perception. Psychological Res., 38, 379-396.

Kaiser, M. K.; and Proffitt, D. R. (1984). The development of sensitivity to causally-relevant dynamic information. Child Develop., 55, 1614-1624.

Kaiser, M. K.; and Proffitt, D. R. (1987a). Observers' sensitivity to dynamic anomalies in collisions. Perception & Psychophysics, 42, 275-280.

Kaiser, M. K.; and Proffitt, D. R. (1987b). Naive mechanics: Erroneous beliefs, but veridical perceptions. Manuscript submitted for publication.

Kaiser, M. K.; Proffitt, D. R.; and Anderson, K. (1985). Judgments of natural and anomalous trajectories in the presence and absence of motion. J. Exp. Psychol.: Learning, Memory, and Cognition, 11, 795-803.

Kubovy, M. (1986). The psychology of perspective and renaissance art. Cambridge, England: Cambridge.

Leeuwenberg, E. (1982). Metrical aspects of patterns and structural information theory. In J. Beck (Ed.), Organization and representation in perception, Hillsdale, NJ: Erlbaum, 57-71.

McCloskey, M. (1983). Intuitive physics. Scientific Amer., 248(4), 122-130.

McCloskey, M.; Caramazza, A.; and Green, B. (1980). Curvilinear motion in the absence of external forces: Naive beliefs about the motion of objects. Science, 210, 1139-1141.

Pritchard, R. M. (1961). Stabilized images on the retina. Scientific Amer., 204, 72-78.

Proffitt, D. R.; and Bertenthal, B. I. (1988). Recovering connectivity from moving point-light displays. In W. N. Martin and J. K. Aggarwal (Eds.), Motion understanding, Hingham MA: Kluwer, 298-328.

Proffitt, D. R.; and Cutting, J. E. (1979). Perceiving the centroid of configurations on a rolling wheel. Perception & Psychophysics, 25, 389-398.

Proffitt, D. R.; and Cutting, J. E. (1980). An invariant for wheel-generated motions and the logic of its determination. Perception, 9, 435-449.

Proffitt, D. R.; Cutting, J. E.; and Stier, D. M. (1979). Perception of wheel-generated motions. J. Exp. Psychol.: Human Percept. Perform., 5, 289-302.

Proffitt, D. R.; Gilden, D. L.; Kaiser, M. K.; and Whelan, S. M. (1988, in press). The effect of configural orientation on perceived trajectory in apparent motion. Perception & Psychophysics.

Rubin, E. (1915). Synsoplevede figurer. Copenhagen: Cyldendalske.

Runeson, S. (1977). On visual perception of dynamic events. Unpublished doctoral dissertation, Univ. Uppsala, Uppsala, Sweden.

Todd, J. T. (1982). Visual information about rigid and nonrigid motion: A geometric analysis. J. Exp. Psychol.: Human Percept. Perform., 8, 238-252.

Todd, J. T.; and Warren, W. H. (1982). Visual perception of relative mass in dynamic events. Perception, 11, 325-335.

Ullman, S. (1979). The interpretation of visual motion. Cambridge, MA: MIT Press.

Wallach, H.; and O'Connell, D. N. (1953). The kinetic depth effect. J. Exp. Psychol., 45, 205-217.

Webb, J. A.; and Aggarwal, J. K. (1982). Structure from motion of rigid and jointed objects. Artificial Intelligence, 19, 107-130.

Wong, E.; and Weisstein, N. (1987). The effects of flicker on the perception of figure and ground. Perception & Psychophysics, 41, 440-448.

Yonas, A.; Craton, L. G.; and Thompson, W. B. (1987). Relative motion: Kinetic information for the order of depth at an edge. Perception & Psychophysics, 41, 53-59.
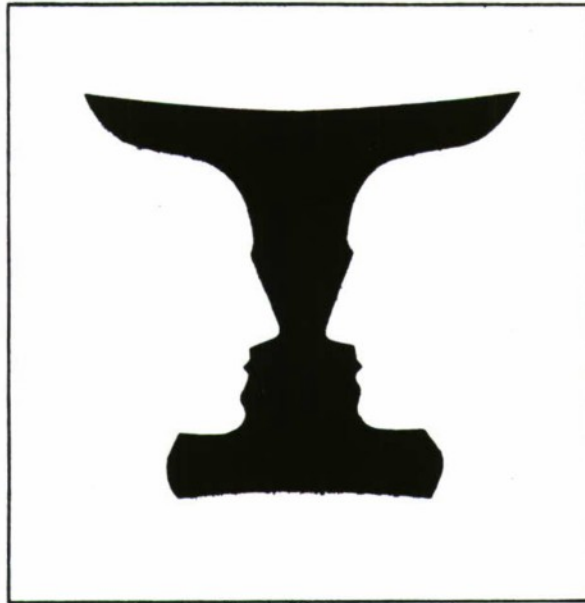
Figure 1.– Rubin's (1915) faces-vase figure.
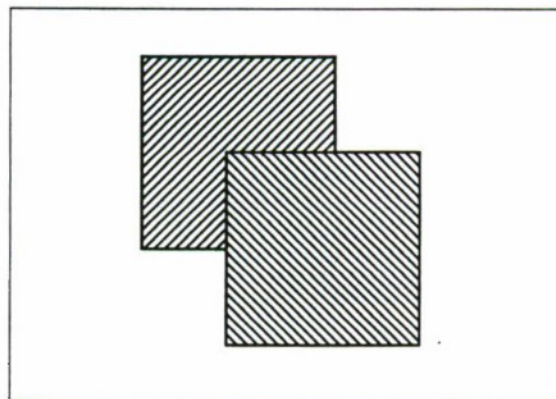


Figure 2.– Two surfaces are depicted. The one to the left appears to partially occlude the surface to the right.
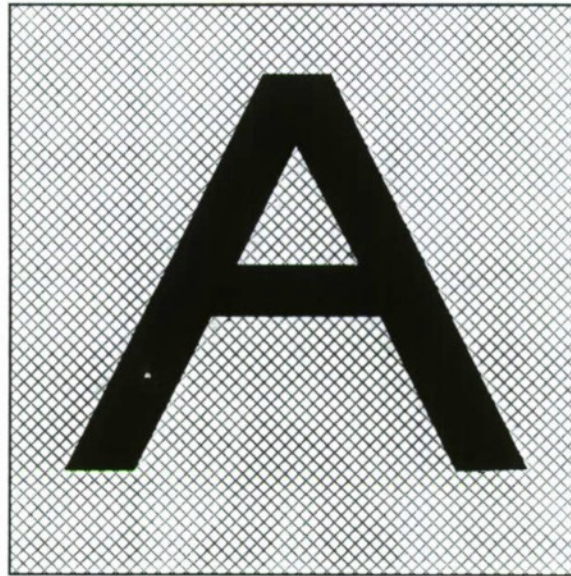
Figure 3.– The familiar figure, A, appears to be in front of the background surface.
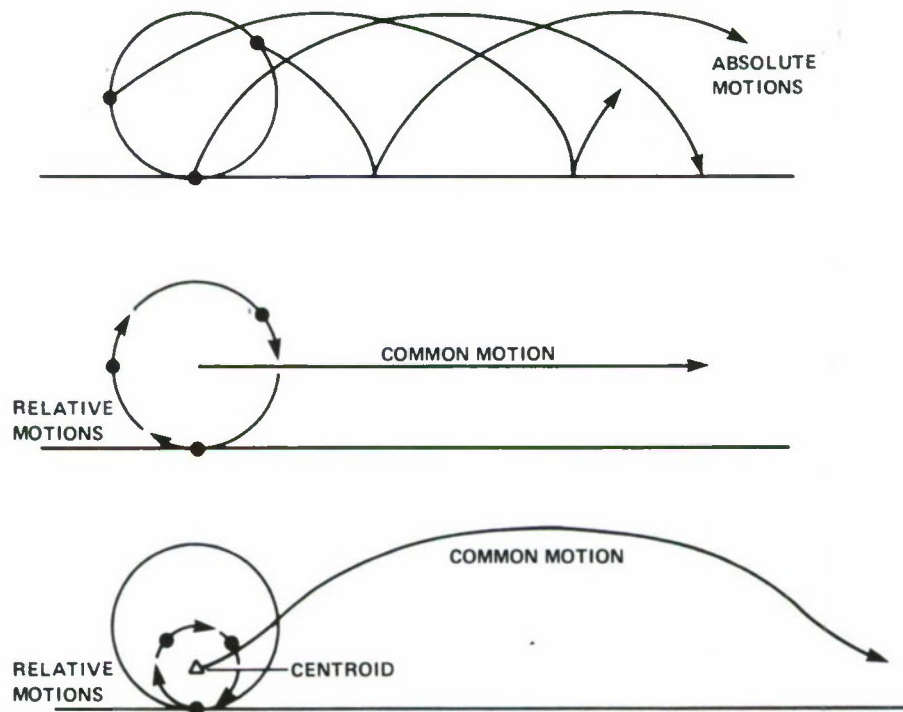


Figure 4.– The top panel depicts the absolute motions of three points on a rolling wheel. The middle panel shows the relative and common motions that are perceived in this event. The bottom panel depicts the perceived motions for three points on a rolling wheel in which the configural centroid of the points does not coincide with the wheel's hub.
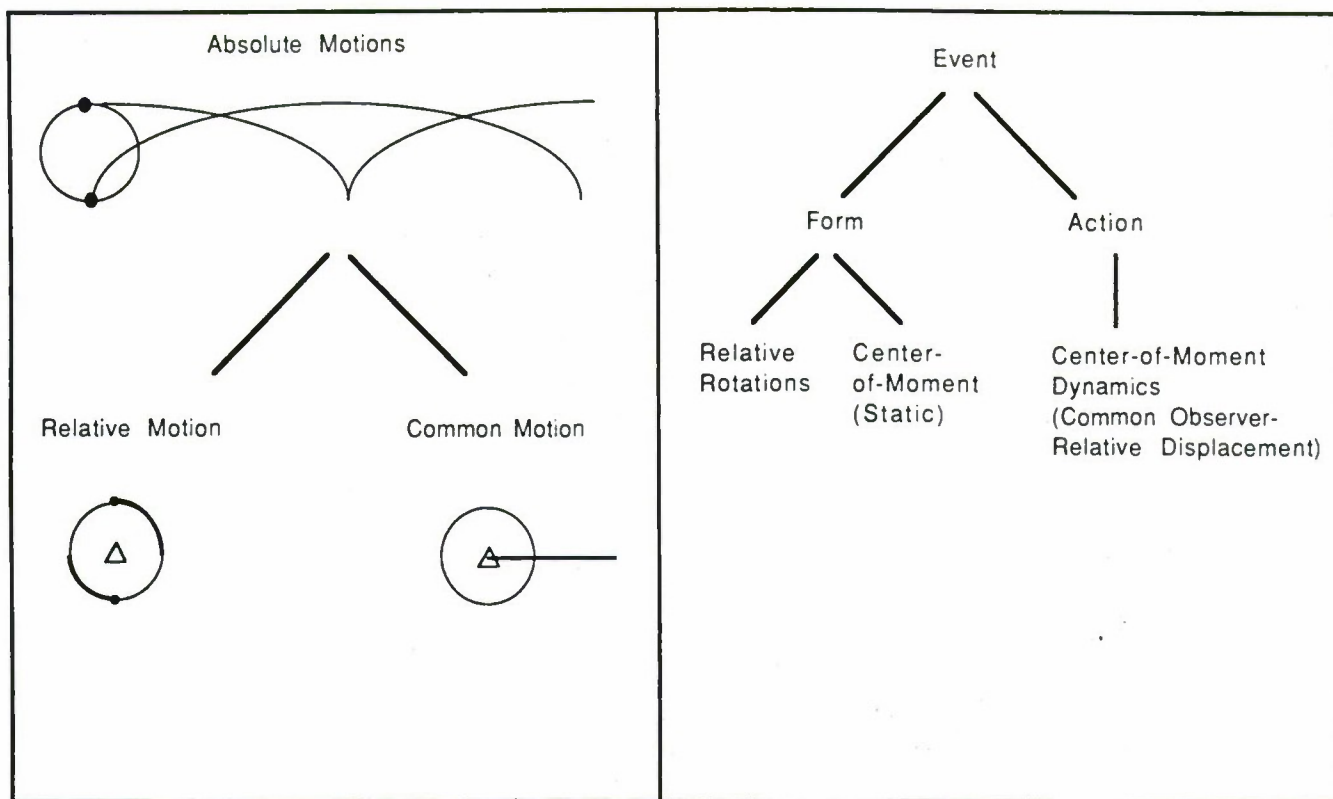
Figure 5.– The perceptual system divides absolute motions into relative and common components. The relative rotations are used in form analysis, whereas the form's common motion defines its observer-relative displacement.
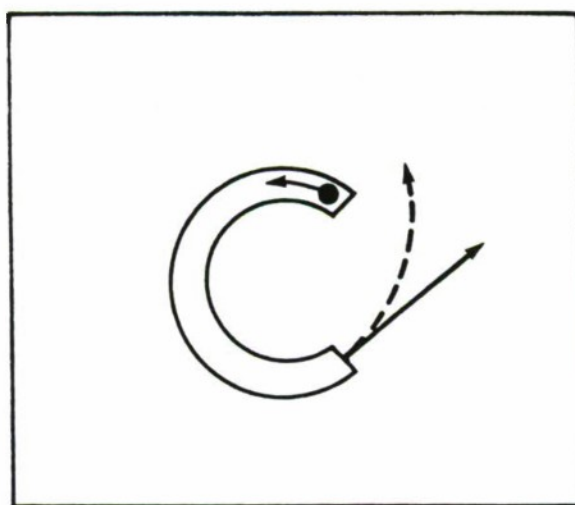


Figure 6.– Depicted is a horizontal C-shaped tube through which a ball is rolled. The two drawn trajectories represent the correct path that the ball takes upon exiting the tube, and a frequently drawn erroneous path.

# VISUAL SLANT UNDERESTIMATION

John A. Perrone
NASA Ames Research Center
Moffett Field, California

Peter Wenderoth
University of Sydney
Sydney, Australia

## SUMMARY

Observers frequently underestimate the in-depth slant of rectangles under reduction conditions. This also occurs for slanted rectangles depicted on a flat display medium. Perrone (1982) provides a model for judged slant based upon properties of the two-dimensional trapezoidal projection of the rectangle. Two important parameters of this model are the angle of convergence of the sides of the trapezoid and the projected length of the trapezoid. We tested this model using a range of stimulus rectangles and found that the model failed to predict some of the major trends in the data. However, when the projected width of the base of the trapezoidal projection was used in the model, instead of the projected length, excellent agreement between the theoretical and obtained slant judgments resulted. The good fit between the experimental data and the new model predictions indicates that perceived slant estimates are highly correlated with specifiable features in the stimulus display.

## INTRODUCTION

Attempts at depicting surfaces slanted in depth on a flat display medium are often hampered by a common perceptual illusion which results in underestimation of the true depth. Surfaces appear to lie closer to the fronto-parallel plane than the perspective projection dictates. This has been a common finding in a wide range of experiments involving slant perception, starting with Gibson's study (1950) on texture gradients (e.g., Clark, Smith and Rabe, 1955; Gruber and Clark, 1956; Smith, 1956; Flock, 1965; Freeman, 1965; Braunstein, 1968; Wenderoth, 1970).

The mode of viewing slanted surfaces under the conditions used in slant perception experiments differs from the way we normally encounter visual slant in our environment (Perrone, 1980). Cutting and Millard (1984) has also questioned the use of slant as a variable in the understanding of surface perception. However, slant underestimation remains an interesting phenomenon because the information is present in the stimulus display for the veridical perception of slant (Perrone, 1982), yet apparently the human visual system does not use that information correctly.

Theories attempting to explain the underestimation are rare. Gogel (1965) applied his "equidistance tendency" theory to slant underestimation effects and Lumsden (1980) speculated that truncation of the visual field by the use of an aperture may be a factor causing underestimation.

Perrone (1980, 1982) has proposed several models of slant perception which attempt to account for the slant underestimation. This paper tests and modifies one of these models. Our aim is to pinpoint the stimulus features used by observers when making visual slant estimates. This would provide useful insights into areas such as spatial orientation, picture perception, and pilot night-landing errors (Perrone, 1984).

## MODEL OF SLANT UNDERESTIMATION

The slant angle $\upsilon$, is obtainable from the two-dimensional projection of the surface onto the retina. (For a technique using perspective lines, see Freeman, 1966; Perrone, 1982.)

The slant angle is found from the two-dimensional variables given in figure 1 using:

$$\theta = \tan^{-1}(\tan \pi/X)f \qquad (1)$$

This equation states that the slant angle, $\theta$, can be derived from the angle of convergence ($\pi$) of the perspective line in the projection, and the distance, X, from the center of the projection out to the perspective line. In equation 1, f is a known constant and it is the arbitrary distance from the eye to the theoretical projection plane used to analyze the array of light reaching the eye.

The convergence angle of perspective lines, $\pi$, can give the slant angle $\theta$ as long as the correct distance X is used. Using a value of X greater than the true value will result in a calculated slant angle less than the actual slant angle, i.e., slant underestimation. Perrone (1980, 1982) proposed a model which suggested that deviation of the perceived straight-ahead direction results in a judgment of slant based on an incorrect value of X.

Two versions of the model have been proposed:

**Model A**. Perrone (1982) suggested that because of the reduced viewing conditions and because of the unusual form of the presenting slant, the observer's perceived straight-ahead direction deviates from the true straight-ahead (fig. 2) and that the visual system uses the length X' (equal to the projected length Y) instead of X.

It is proposed that the visual system is attempting to measure the change in width over a square area of the projection plane, determined by Y, but because there are no perspective lines a distance X' out from c', the outside edge of the rectangle is used instead. When X' is substituted into equation (1) instead of X, the equation for perceived slant becomes $\beta = \tan^{-1}(\tan\pi/X')f$. However, in order to use this equation for predicting perceived slant, we need to replace the two-dimensional variables ($\pi$ and X') with the three-dimensional parameters of the stimulus situation. This gives the following equation for perceived slant:

$$\beta = \tan^{-1}\left[\frac{W \sin \theta \, (D^2 - L^2 \sin^2 \theta)}{4 L D^2 \cos^2 \theta}\right] \qquad (2)$$

$\theta$ = actual slant
W = actual width of rectangle

L  = half the total length of rectangle
D  = distance from eye to center of rotation

To date, Perrone (1982) has shown how this sort of analysis provides acceptable fits to data collected by others (e.g., Clark, Smith, and Rabe, 1955; Smith, 1956), but these studies were designed to investigate other aspects of slant perception and so did not involve direct manipulation of the variables integral to the model.

One problem with this version of the model is that it predicts that slant *overestimation* will occur when the projected height of the test rectangle (Y) becomes less than the projected half-width at the axis of rotation (X). However, there have been no published accounts of slant overestimation occurring, but this may simply be because nobody has used test rectangles with the appropriate length-to-width ratio.

**Model B**. (Modified version of Model A). This version proposes that the total base width of the rectangle ($X_b$) is used in the evaluation of the slant angle instead of X. This new form of the model can be interpreted as saying that the observers are basing their slant estimates on the convergence angle, $\pi$, of perspective lines which they believe to be twice the true distance out from the center. It may be that it is a difficult and unnatural task for the observer to judge the slant of a surface which is centered on the median plane of the eye. It is easier if we have a side view or at least a more oblique view of the slanted surface. The observers may resort to making their judgments on the basis that they have a more extreme or displaced viewpoint than is in fact the case. Their interpretation of the slant of the rectangle may be based on an assumed view of the rectangle which is displaced or rotated relative to its true position.

When this error is combined with the proposed deviation of the perceived straight-ahead (Perrone 1982), the result may be the erroneous use of the total base width of the projected trapezoid rather than the correct half-width at the axis of rotation. When the total projected base width of a slanted rectangle is used to estimate theta from equation 1, the predicted perceived slant angle is found using

$$\beta = \tan^{-1} \left[ \frac{\tan \theta \, (D - L \sin \theta)}{2 \, D} \right] \tag{3}$$

$\theta$  = actual slant
L  = half the total length of rectangle
D  = distance from eye to center of rotation

## TESTING THE MODEL

An experiment was designed to verify which of the two cases (equation 2 or equation 3) best models the data from human observers in the slant perception task. If it can be established that specific features of the stimulus display are being used in the slant estimation process, then the more difficult task of discovering why these particular variables are being used can be attempted. The model provides a means of narrowing down the choice of possible variables and the combination in which they are used.

## Experiment

The stimuli were computer-generated two-dimensional perspective representations of rectangular outline figures, presented on a CRT and viewed monocularly through an aperture. These figures represented rectangles measuring 25 cm wide with the following lengths: 50 cm (condition 1), 25 cm (condition 2), and 15 cm (condition 3). These were depicted to be at a distance of 57 cm from the subject's eye and slanted backwards away from the observer by varying angles of slant. The actual slant angles used were 20° 40°, 60° and 80° measured from the vertical.

The subject reproduced the judged slant of the rectangle on a response device which was located 90° to the right and positioned at eye level. The response device consisted of a thin black line inscribed on a clear plexiglass strip which was mounted on a circular white metal disk 23 cm in diameter. Vertical and horizontal black lines were drawn on the disk to provide anchor points (Wenderoth, 1970). Subjects were 10 paid volunteers, naive as to the aims of the experiment.

## Predictions

If Model A is correct, then the slant estimates for the three different conditions should lie along three distinct curves given in figure 3a. For some of the stimulus conditions, the subjects should judge the rectangle to be slanted farther back from the fronto-parallel plane than the true position (slant *overestimation*). This corresponds to any region of the curves which lies above the dotted line in figure 3a. If a Model B is correct, the slant estimates for all three conditions should all lie on approximately the same curve of the shape shown in figure 3b. No slant overestimation should occur.

## Results

The data from the 10 subjects have been plotted in figure 4 along with the predictions from Model B. For the case in which a tall narrow rectangle was used (Condition 1), the results are similar to those obtained in past slant perception experiments which used rectangles with a length-to-width ratio greater than one, (e.g., Smith, 1956). For this condition, both Model A and B give reasonable predictions for the smaller test angles (see C1 predictions in fig. 3a). However, for the remaining conditions, the data depart greatly from the Model A predictions and none of the predicted overestimation of slant occurred.

The mean absolute error between the Model A predictions and the data over the three conditions was 13.9°, (sd = 8.1). For Model B, on the other hand, the mean absolute error was only 2.6°, (sd = 1.9). The mean absolute errors from Model A are significantly greater than those from Model B, (t = 4.5, p < 0.05, 22df) and represent a worse fit between the model predictions and data.

## CONCLUSIONS

Slant underestimation Model A (Perrone 1982) incorrectly predicts overestimation to occur for rectangles which have a projected length less than half of the base width. In fact, the influence of

the projected length of the rectangle on slant judgments is minimal. However, Model B provides an excellent fit between the experimental data and the predictions. These predictions are based on measurable features of the experimental configuration. There are no free parameters. Model B states that the total projected base width of the rectangle is used instead of half the projected width at the axis of rotation. Two parameters of the two-dimensional projection are important in the slant estimation process: (1) the angle of convergence of perspective lines and (2) the distance of the perspective lines from the center of the projection. The success of Model B suggests the human observers make errors in slant estimates because they misperceive this second parameter.

The question remains as to why human observers use "incorrect" features of the stimulus in their assessment of the slant angle. It has been shown that the correct slant angle is obtainable from the appropriate use of the variables given in equation 1. These variables are known to be present in the two-dimensional stimulus reaching the observer's eye. The experimental data are consistent with the proposal that the total base width of the trapezoidal projection is used instead of half the projected width at the axis of rotation. However, it does not shed any light as to why this should be the case.

Further research is required before we can conclude the actual mechanisms used by the human visual system in making slant estimates. In the meantime, sufficient evidence exists to conclude that slant judgments by an observer are highly correlated with specific measurable features in the two-dimensional array of light reaching the observer's eye. The slant estimates exhibit a large amount of error and often greatly underestimate the true slant angle. This paper shows that such errors cannot be attributed to the fact that insufficient information exists in the stimulus for veridical slant judgments. The information is available, but is incorrectly used.

# REFERENCES

Braunstein, M. L.: Motion and texture as sources of slant information. J Exper. Psychol., vol. 8, 1968, pp. 584-590.

Clark, W. C.; Smith, A. H.; and Rabe, A.: Retinal gradients of outline as a stimulus for slant. Canadian J. Psychol., vol. 9, 1955, pp. 247-253.

Cutting, J. C.; and Millard, R. T.: Three gradients and the perception of flat and curved surfaces. J. Exper. Psychol.: General, vol. 113, 1984, pp. 198-216.

Flock, H. R.: Optical texture and linear perspective as stimuli for slant perception. Psychol. Rev., vol. 72, 1965, pp. 505-514.

Freeman, R. B., Jr.: Ecological optics and visual slant. Pyschol. Rev., vol. 72, 1965, pp. 501-504.

Freeman, R. B., Jr.: Function of cues in the perceptual learning of visual slant. Psychol. Monogr.: Gen. Appl., vol. 80, 1966, (2) whole no. 610.

Gibson, J. J.: The perception of surfaces. Amer. J. Psychol., vol. 63, 1950, pp. 367-384.

Gogel, W. C.: Equidistance tendency and its consequences. Psychol. Bull., vol. 64, 1965, pp. 153-163.

Gruber, H. E.; and Clark, W. C.: Perception of slanted surfaces. Percept. Motor Skills, vol. 16, 1956, pp. 97-106.

Lumsden, E. A.: Problems of magnification and minification: An explanation of the distortions of distance, slant, shape and velocity. In M. A. Hagen (Ed.) The Perception of Pictures, vol. 1, 1980, (pp. 91-134). New York: Academic Press.

Perrone, J. A.: Slant underestimation: A model based on the size of the viewing aperture. Perception, vol. 9, 1980, pp. 285-302.

Perrone, J. A.: Visual slant underestimation: A general model. Perception, vol. 11, 1982, pp. 641-654.

Perrone, J. A.: Visual slant misperception and the 'black-hole' landing situation. Av. Space Environ. Medicine, vol. 55, 1984, pp. 1020-5.

Smith, A. H. Gradients of outline convergence and distortion as stimuli for slant. Canadian J. Psychol., vol. 10, 1956, pp. 211-218.

Wenderoth, P. M.: A visual spatial after-effect of surface slant. Amer. J. Pyschol., vol. 83, 1970, pp. 576-590.
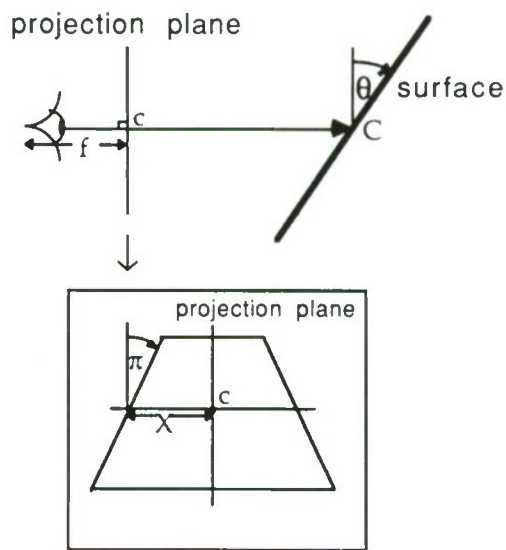
Figure 1.– The two-dimensional information reaching the eye is analyzed on a theoretical projection plane an arbitrary distance f from the eye. All measurements on the projection plane are made within the plane of the page.
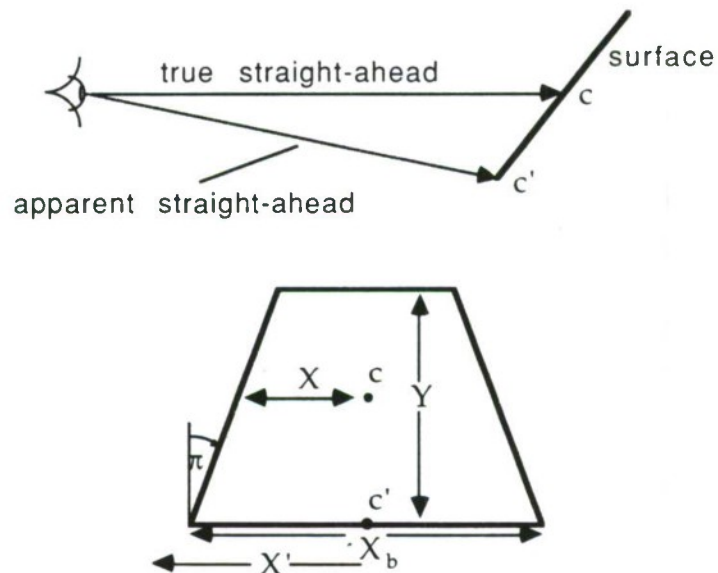


Figure 2.– Deviation of the perceived straight-ahead results in the analysis being carried out about c' instead of c. Model A states that the length X' (equal to Y) is used instead of X. Model B proposes that $X_b$ is used instead of X.
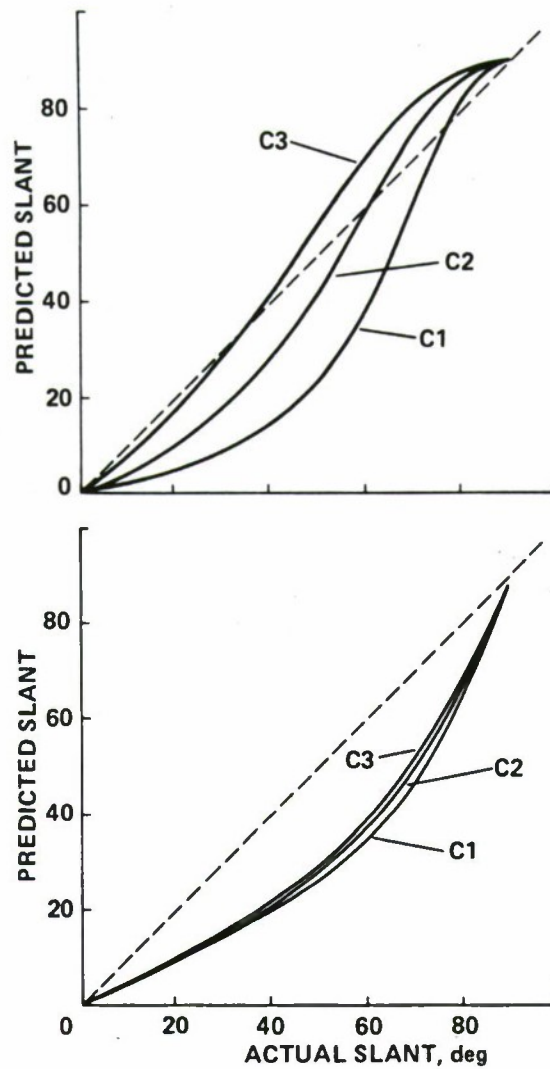
Figure 3.– Plots showing (a) predictions from Model A for each of the three experimental conditions and (b) predicted slant versus actual slant for Model B. No slant overestimation is predicted to occur.
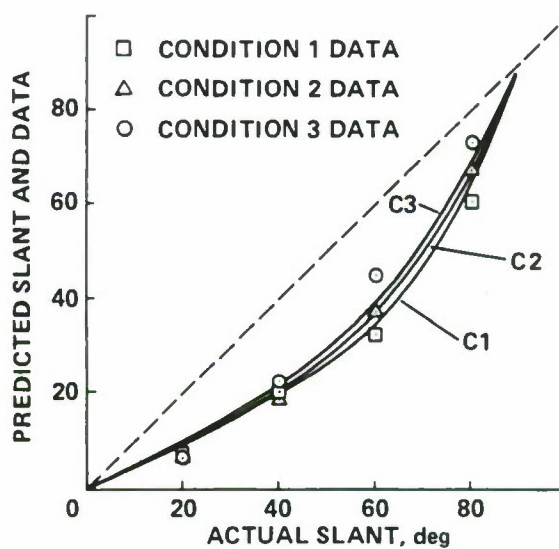
Figure 4.– Data are plotted from conditions 1, 2, and 3 along with predictions from Model B. Error bars have been omitted for clarity, but the largest standard error was 4.5° for the 80° slant angle.

# THE PHOTO-COLORIMETRIC SPACE AS A MEDIUM FOR THE REPRESENTATION OF SPATIAL DATA

K. Friedrich Kraiss and Heino Widdel
Research Institute for Human Engineering
D-5307 Wachtberg-Werthhoven
Federal Republic of Germany

## SUMMARY

Spatial displays and instruments are usually used in the context of vehicle guidance, but it is hard to find applicable spatial formats in information retrieval and interaction systems. This paper discusses human interaction with spatial data structures and the applicability of the CIE color space to improve dialogue transparency. A proposal is made to use the color space to code spatially represented data. The semantic distances of the categories of dialogue structures or, more general, of database structures, are determined empirically. Subsequently the distances are transformed and depicted into the color space. The concept is demonstrated for a car diagnosis system, where the category "cooling system" could, e.g., be coded in blue, the category "ignition system" in red. Hereby a correspondence between color and semantic distances is achieved. Subcategories can be coded as luminance differences within the color space.

## INTRODUCTION

The increasing dissemination of information technology as well as the expanding complexity of computer systems require user-friendly interaction techniques. One design goal of high relevance in the context of user friendliness is the transparency of system functions. In general, transparency is defined as a well-structured, consistent, and comprehensible appearance of the system for its users (Widdel and Kaster, 1986). One way to reach transparency consists of the design of a suitable menu structure. Especially for occasional and untrained users of computer systems a menu-based dialogue is of great advantage.

The designer of dialogues has to analyze the characteristics of the expected user group in order to adapt the dialogue interface to the mental model of the users. Knowledge of specific cognitive human behavior must guide the design of human-computer interaction in general, and of dialogue structures in particular.

A systematic or intuitive transfer of this basic knowledge of cognitive functions leads to iconic visualization of information in human-computer interaction. By presenting user commands and system information in iconic form, as pictures or three-dimensional presentations, better use is made of human visual capabilities.

# GRAPHICAL DESIGN OF DIALOGUE STRUCTURE

The proposals made in this paper aim at further improving the graphical presentation of dialogue structures by considering three-dimensional concepts. This expands earlier work on dialogue design performed by Kaster and Widdel (1987). In comparing various dialogue designs, they used a conventional menu as given in figure 1a showing a menu with a set of five available choices. It includes title, menu options, selection codes, and the user query. Alternatively, they displayed the hierarchical organization of the dialogue structure as a picture. It encloses the total range of functions or menus offered in the dialogue. This picture is presented in figure 1b. The hypothesis underlying this experimental setup postulated that an interface design using a graphic conceptual model can facilitate the formation of an appropriate mental model of the interactive computer system (Bennett, Parasuraman, and Howard, 1984). The experiments of Kaster and Widdel confirmed this hypothesis and demonstrated that naive computer users can successfully run the dialogue with this interface.

The dialogue presented in figure 1 was used for experimental reasons and restricted to a relatively low complexity; real applications require much more complex dialogue structures. In terms of user-friendliness, research activities are focused on the breadth and depth as two relevant dimensions of dialogue complexity. Intensive and detailed discussions and investigations (MacGregor and Lee, 1987; Paap and Roske-Hofstrand, 1986) expand this problem area from the pure interaction field to the more general perspective of searching data bases.

High-resolution, direct-manipulation interfaces have been monochrome for a long time for technical reasons. As these restrictions are no longer valid, it is about time to consider reasonable applications of color. Distinct overviews of human factors knowledge about the use of color in visual displays is given by Davidoff (1987), Murch (1985), and van Nes (1986). In the context of this paper it will be of particular interest to show in which way color can be used to convey information about spatial structures instead of or in addition to 3-D graphics. For this purpose the colormetrics and psychometrics of color will be discussed in the next section.

## COLORMETRICS AND PSYCHOMETRICS OF THE COLOR SPACE

Color can be defined by chromaticity and luminance; together they establish the photocolorimetric space (subsequently more simply called "color-space") as depicted in figure 2. The base plane described by the coordinates u' and v' defines the chromaticity of a color, while the third axis L gives the luminance (CIE, 1977). The luminance achievable with a standard TV monitor varies between 20 and 200 $cd/m^2$ depending on the color. Typical chromaticity coordinates are 0.42/0.54 for red, 0.12/0.57 for green, and 0.16/0.18 for blue. With these data the solid depicted in figure 2 roughly describes the color space available on commercial monitors. A color of particular chromaticity and luminance corresponds to a point in this color space (Kaster, Kraiss, and Küttelwesch, 1985).

The number of distinguishable points in the color space can be estimated from the number of just noticeable differences in chromaticity ($jnd_C$) and luminance ($jnd_L$).

The number of just noticeable luminance differences ($jnd_L$) is defined by the available luminance range and by the size of a threshold step. For the purposes of this paper we make use of a threshold contrast $C_L = 1.05$. This results in (Galves and Brun, 1975):

$$jnd_L = \log 1.05 = 0.021 \qquad (1)$$

For comfortable discernibility, a value seven times larger usually is applied, i.e.:

$$jnd_L* = 7 \times jnd_L = 0.15 \qquad (2)$$

According to (1) a luminance range from 10 to 100 $cd/m^2$ can accommodate

$$(\log 100 - \log 10)/0.021 = 47.6 \ jnd_L\text{'s}.$$

For the threshold chromaticity difference $jnd_C$ Galves and Brun (1975) proposes a value of 0.00384 as the smallest color difference the eye can discern. Again, for practical purposes it is common practice to use a value seven times larger than the threshold for easy discernibility

$$jnd_C* = 7 \times jnd_C = 0.027 \qquad (3)$$

As an example we calculate with the numbers given above the distance between red and blue to be $(\Delta u'^2 + \Delta v'^2)^{1/2} = 0.354$. Hence, a total of $0.354/0.00384 = 92 \ jnd_C$'s can be accommodated between these two colors. For simultaneous variations in luminance and chromaticity the number of discernible steps is determined by

$$jnd_{CL} = (jnd_C + jnd_L)^{1/2} \qquad (4)$$

The photo-colorimetric space depicted in figure 2 offers ample opportunity for the composition of chromaticity/luminance trajectories. With respect to limited space only two representative examples are presented here. Tables 1 and 2 give their u',v',L-coordinates together with the number of jnd's contained in a particular trajectory (see also the corresponding figs. 3 and 4).

From previous experience in experiments with color-coded sensor data, it appears that observers can make a rather accurate estimate of distances in the color space (Kraiss and Küttelwesch, 1984). The number of absolutely discriminable states in the color space is, of course, much less than the number of jnd's. For chromaticity usually 6 to 9 and for luminance usually 6 values can be distinguished with sufficient reliability.

## SEMANTICS AND COLOR SPACE

Any structure of a dialogue or database has a semantic system of categories underlying the organization. For example, a *car diagnosis system* contains the categories *electric system, suspension system, ignition system, cooling system, fuel system, and gear system* with appropriate subcategories on lower levels. The semantic distances of these categories can be determined empirically using multivariate methods of similarity scaling. The resulting similarity ratings establish a spatial structure, or semantic net, that may be used to build menu structures. Roske-Hofstrand and

Paap (1986) used this procedure to define menu organizations matched to the semantic net of experts for a cockpit information system.

Semantic distances can be depicted as chromaticity differences in the color space (fig. 5). In our example the categories *ignition* (C) and *cooling* (D) are separated by a long semantic distance which finds its equivalent in the long distance from red to blue. The categories *electric* (A) and *gear* (F), having a shorter semantic distance, are assigned to the colors green and cyan.

In selecting colors for menu options or categories, the psychology of color perception must be taken into account. Besides the correspondence of distances of both spaces, the problem of association between a category and a color arises, i.e., should category D be colored blue and category C red or vice versa. This problem can be solved empirically; sometimes appointments are predefined by tradition. While the association of blue with a cooling system and of red with an ignition system is evident, this is not the case for yellow (suspension system), green (electric system), cyan (gear system), and violet (fuel system).

Luminance as the third dimension of the color space may be used for coding the lower hierarchical levels of a menu structure or database (fig. 5) while retaining the chromaticity of the top-level category. Each category coded by a specific chromaticity is varying luminance with corresponding lower levels. In figure 5 the *cooling system* (D1) on the highest level may have a luminance of 24 $cd/m^2$. On the second level the *cooling system* could have, among others, the subcategories *water cooling* and *air cooling* (D2n). They will be assigned the same chromaticity coordinates, but on the second luminance level of 15 $cd/m^2$. On the third level a subcategory of water cooling could be water supply (D3nn) with a possible luminance of 5 $cd/m^2$.

Another possible application of color for the orientation in a multidimensional data space is proposed by Korfhage (1986). He describes a browser concept for navigating through a database by visual support. Browsing is defined as a dynamic search through an information resource, with no specific goal initially in mind. He models a set of documents as an n-dimensional space and simulates browsing by a loosely directed traversal of this space. Making use of the Doppler effect, documents far ahead of the actual search position were color-coded with blue; those far behind were color-coded with red. The document nearest to the user's plane is represented in yellow; transition color to blue is green and to red is orange.

## CONCLUSIONS

A concept for the use of color to convey spatial information at the user interface was discussed. It was suggested that the color space can be used to represent spatially distributed or hierarchically organized data. This implies that an operator can form a corresponding mental color space model that enables him to associate chromaticity/luminance distances to geometric distances. Earlier experiments with color-coded sensor data suggest that this is possible. In an example a possible application of this concept to a car diagnosis database was described.

# REFERENCES

Bennett, K. B., Parasuraman, R., and Howard, J. H. (1984): Mental models in interface design: the role of graphic displays. In: A. Mital (ed.) Trends in ergonomics/human factors I (pp. 113-118). Elsevier Science Publishers: Amsterdam.

CIE (1977): Official recommendations on uniform color spaces, color difference equations, and metric color terms. (Suppl. No. 2, Publ. No. 15 to Colorimetry). Bureau Central CIE: Paris.

Davidoff, J. (1987): The role of color in visual displays. In: D. J. Osborne (ed.) International Reviews of Ergonomics (pp. 21-42). Taylor & Francis: London.

Galves, J. P. and Brun , J. (1975): Color and brightness requirements for cockpit displays: proposal to evaluate their characteristics, AGARD-CP-167, 61-68.

Kaster, J. and Widdel, H. (1987): Effect of visual presentation of different dialogue structures on human-computer interaction. In: B. Knave and P.-G. Widebäck (eds.) Work with display units 86 (pp. 822-830). Elsevier Science Publishers: Amsterdam.

Kaster, J., Kraiss, K. F., and Küttelwesch, H. (1985): Ein Verfahren zur Grafisch-Interaktiven Definition von Farbvalenzen. In: H.-J. Bullinger (Hrsg.) Software-Ergonomie '85. (pp. 396-405). Teubner Verlag: Stuttgart.

Kraiss, K. F. and Küttelwesch, H. (1984): Color and grey scale in sonar displays. In: Society for Information Display, Symposium, June, San Francisco, 68-71.

Korfhage, R. R. (1986): Browser–a concept for visual navigation of a database. In: IEEE Computer Society workshop on visual languages (pp. 143-148). IEEE Computer Society: Dallas, Tex.

MacGregor, J. N. and Lee, E. S. (1987): Performance and preference in videotex menu retrieval: a review of the empirical literature. Behaviour Information Technol., 6, 43-68.

Murch, G. M. (1985): Colour graphics–blessing or ballyhoo? Computer Graphics Forum, 4, 127-135.

Paap, K. R. and Roske-Hofstrand, R. J. (1986): The optimal number of menu options per panel. Human Factors, 28, 377-385.

Roske-Hofstrand, R. J. and Paap, K. R. (1986): Cognitive networks as a guide to menu organisation: An application in the automated cockpit. Ergonomics, 29, 1301-1311.

van Nes, F. L. (1986): Space, colour and typography on visual display terminals. Behaviour Information Technol., 5, 99-118.

Widdel, H. and Kaster, J. (1986): Transparency of a dialogue through pictorial presentation of the dialogue structure. In: H.-P. Willumeit (ed.) <u>Human decision making and manual control</u> (pp. 135-143). Elsevier Science Publishers: Amsterdam.

Table 1.– Chromaticity/luminance trajectory covering 249 jnd's. Presented are color scale, color space coordinates, and jnd's.

| Reference | jnd's$_{CL}$ | u | v | L cd / m$^2$ |
|---|---|---|---|---|
| 1 | | 0,19 | 0,31 | 1 |
| | 48 | | | |
| 2 | | 0,16 | 0,12 | 2 |
| | 41 | | | |
| 3 | | 0,28 | 0,22 | 5 |
| | 51 | | | |
| 4 | | 0,42 | 0,36 | 12 |
| | 59 | | | |
| 5 | | 0,19 | 0,37 | 28 |
| | 22 | | | |
| 6 | | 0,12 | 0,38 | 64 |
| | 28 | | | |
| 7 | | 0,19 | 0,31 | 150 |

$\Sigma = 249$

Table 2.– Luminance scales for 6 chromaticities applicable to menu design.  Presented are color scale, color space coordinates, and jnd's.

| Reference | jnd's$_{CL}$ | u | v | L cd / m$^2$ |
|---|---|---|---|---|
| 1 | | | | 1 |
| | 67 | 0,16 | 0,12 | |
| 2 | | | | 27 |
| 3 | | | | 6 |
| | 65 | 0,13 | 0,30 | |
| 4 | | | | 150 |
| 5 | | | | 6 |
| | 65 | 0,12 | 0,38 | |
| 6 | | | | 150 |
| 7 | | | | 6 |
| | 65 | 0,19 | 0,37 | |
| 8 | | | | 150 |
| 9 | | | | 2 |
| | 62 | 0,42 | 0,36 | |
| 10 | | | | 42 |
| 11 | | | | 3 |
| | 67 | 0,28 | 0,22 | |
| 12 | | | | 80 |

$\Sigma = 391$
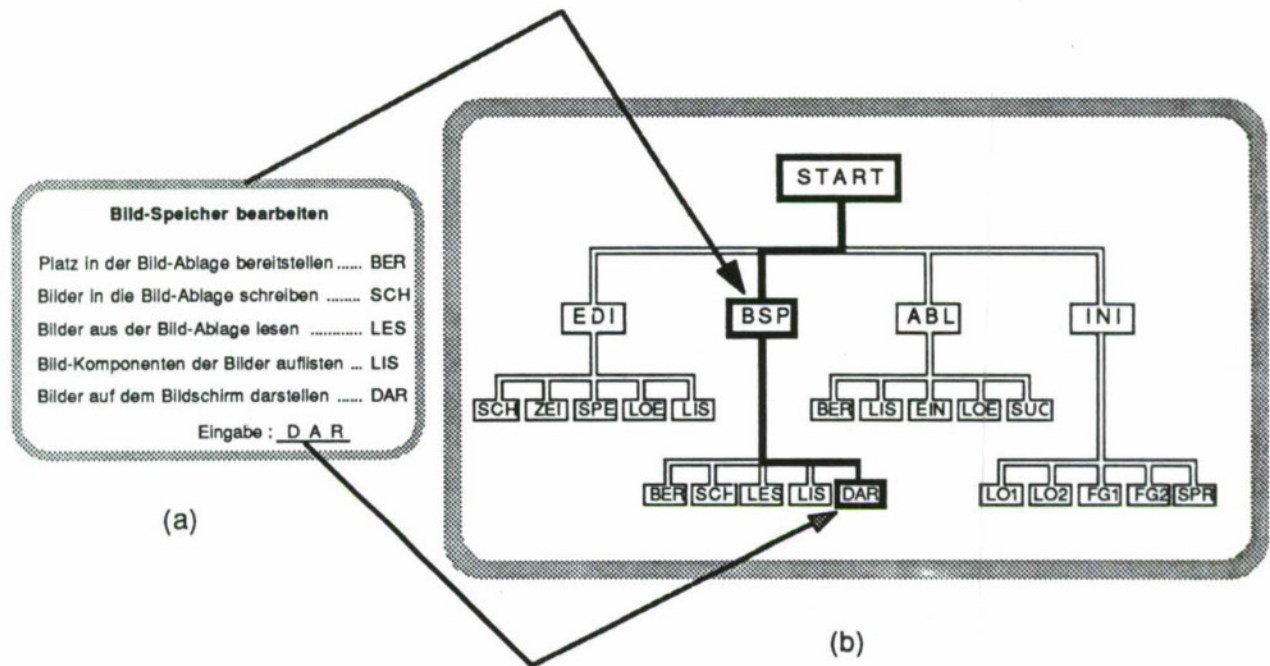
Figure 1.– Textual menu (a) and corresponding picture of the entire dialogue structure (b) (Widdel and Kaster, 1986).
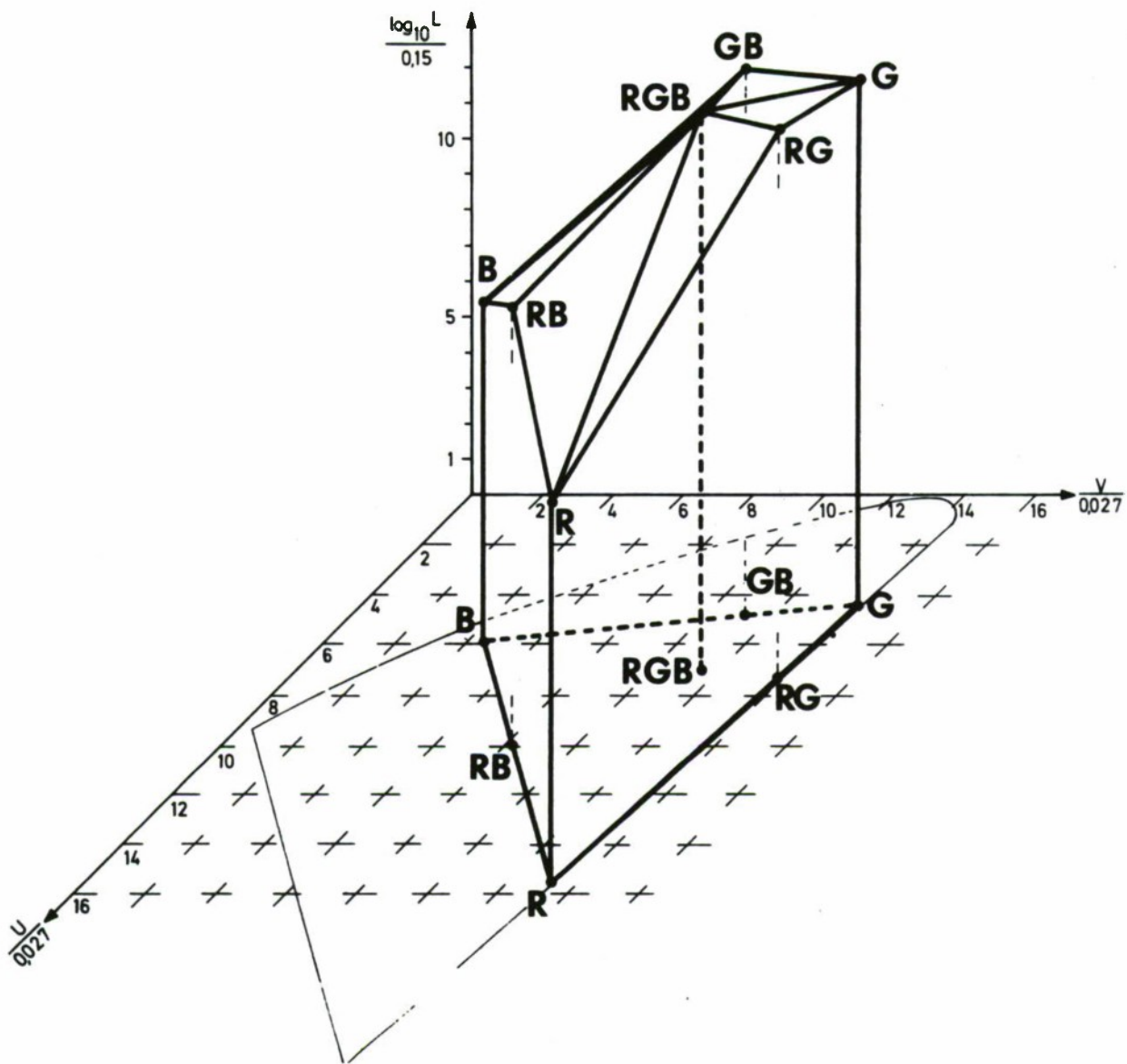
Figure 2.– The photo-colorimetric space with metrics of Galves and Brun (1975). The axes are scaled to just noticeable differences (jnd's).
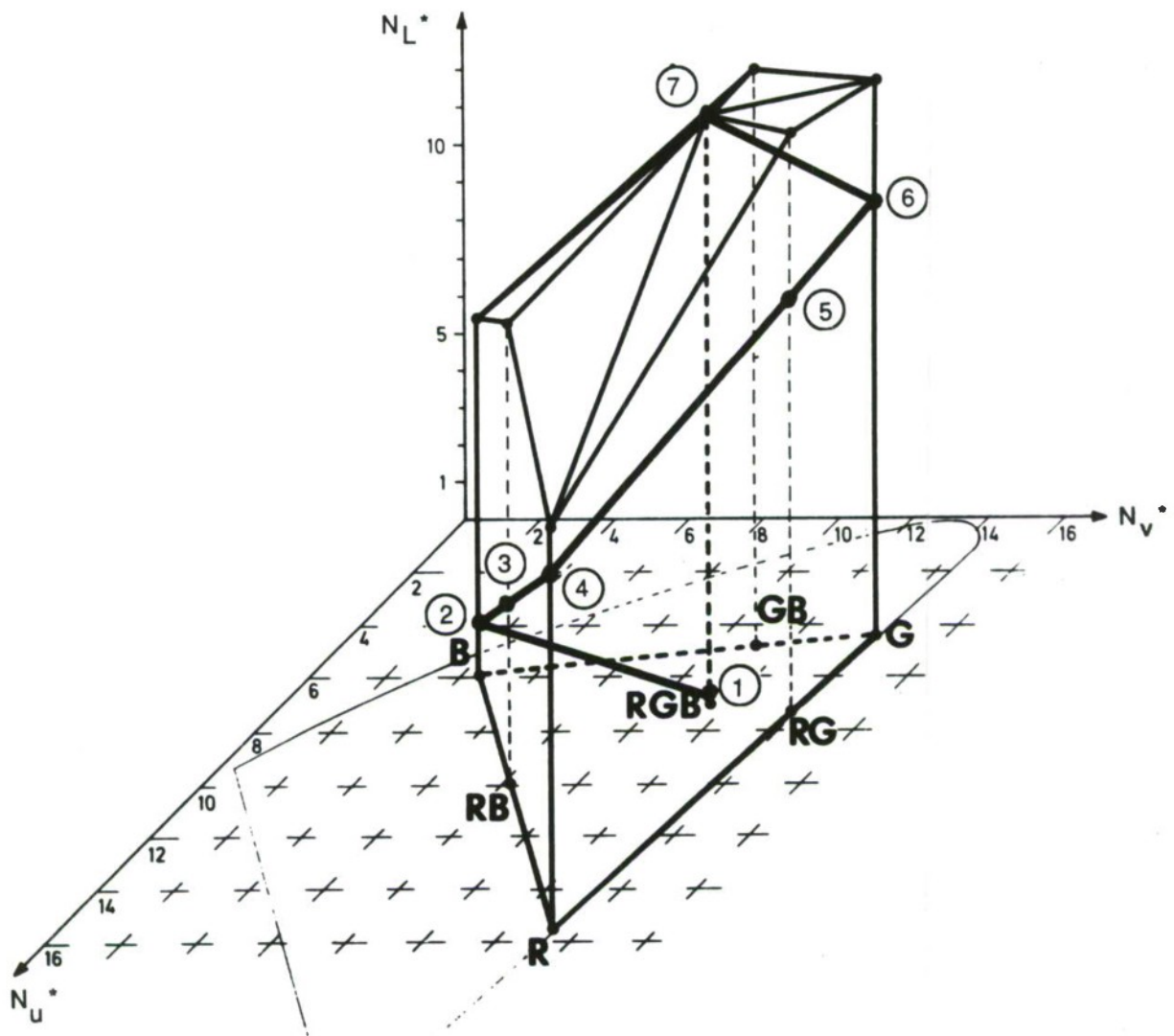
Figure 3.– Color space trajectory corresponding to the values given in table 1.

Figure 4.– Color space trajectory corresponding to the values given in table 2.

(a)

A = electric    D = cooling
B = suspension  E = fuel
C = Ignition    F = gear

A = green   (electric)    D = blue    (cooling)
B = yellow  (suspension)  E = violet  (fuel)
C = red     (ignition)    F = cyan    (gear)

(c)

1 = high luminance
2 = medium luminance
3 = low luminance
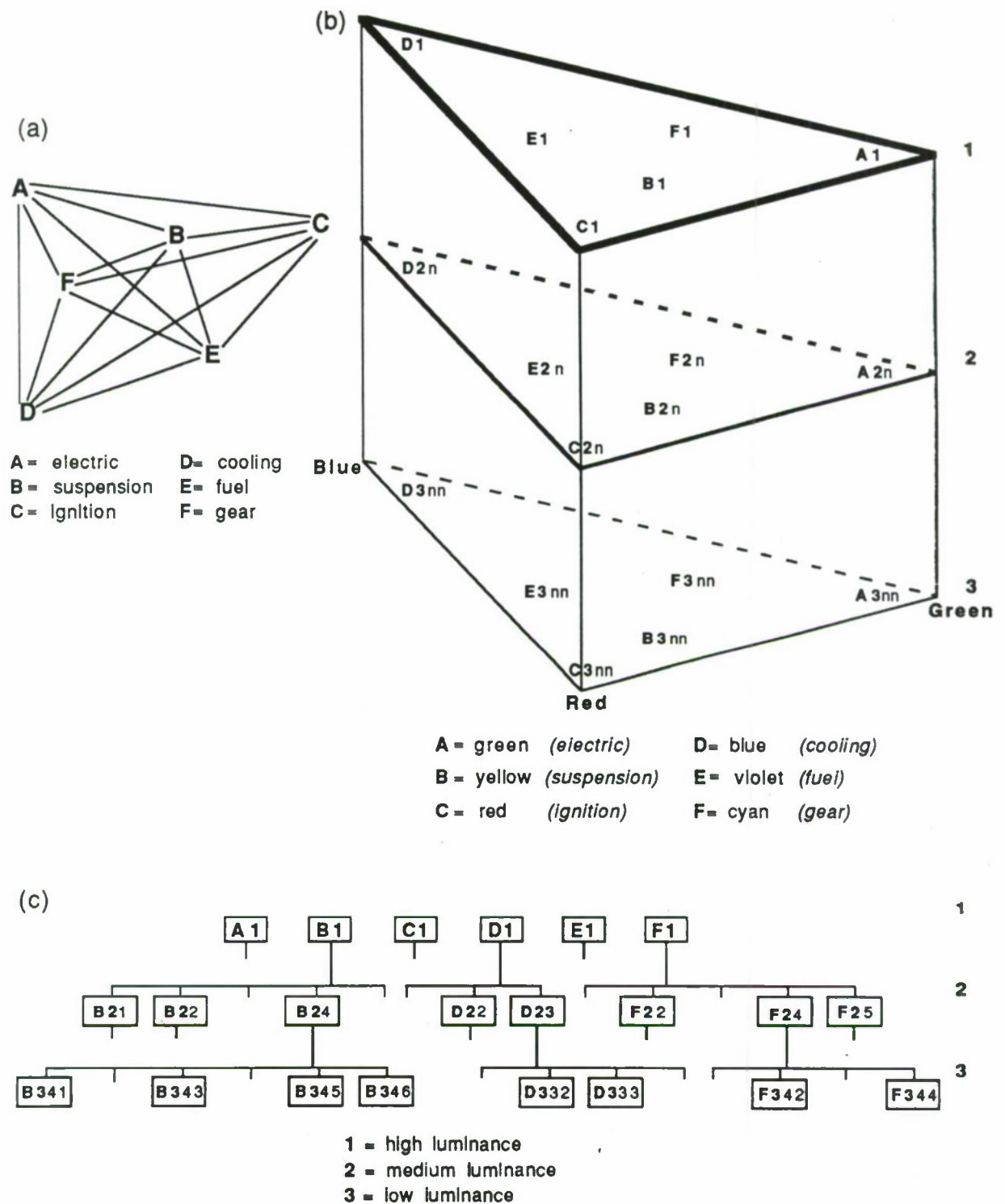
Figure 5.– (a) Fictitious net of semantic distances for categories in a car diagnosis system. (b) The semantic net from (a) mapped onto the chromaticity plane. Three luminance levels are used to accommodate hierarchy subitems (see fig. 2). (c) Two-dimensional dialogue structure with additional chromaticity/luminance assignments to visualize semantic distances and hierarchy levels.

# SPATIAL ORIENTATION

# SPATIAL VISION WITHIN EGOCENTRIC AND EXOCENTRIC FRAMES OF REFERENCE

Ian P. Howard

Human Performance Laboratory, Institute for Space and Terrestrial Science
York University, Toronto, Ontario

## 1. INTRODUCTION

Our ability to perceive a stable visual world and judge the directions, orientations and movements of visual objects is remarkable given that the images of objects may move on the retina, the eyes may move in the head, the head may move on the body, and the body may move in space. An understanding of the mechanisms involved requires that definitions of relevant coordinate systems be as precise as possible. An *egocentric* frame of reference is defined with respect to some part of the observer. When both the object being judged and the reference frame are parts of the body, we have a *proprioceptive* task. If the object being judged is external to the body, its position, orientation and movement may be judged with respect to any of three principal egocentric coordinate systems, an *oculocentric* frame associated with the eye, a *headcentric* frame associated with the head and a *bodycentric* frame associated with the torso. A reference frame external to the body is an *exocentric* frame. In an exocentric task the object being judged may be part of the body, as when a person points north, or it may be external to the body, as when a person judges the direction of one object with respect to another. In addition there are reference frames which combine egocentric and exocentric elements. For instance, when we say that an object is north of us, we use our own body as the origin of a directional scale which is also anchored to the world. The same is true when a person says that something is above the head. Such frames may be referred to as *heterocentric* frames of reference. These various frames of reference are listed in table 1 together with examples of judgments of each type.

Polar coordinates based on meridional angles and angles of eccentricity are commonly used for the objective specification of the oculocentric position of a visual object. The subjective registration of the oculocentric position of an object depends on the local sign mechanism of the visual system. This is the mechanism whereby, for a given position of the eye, each region of the visual field has a unique (one-to-one) and stable mapping onto the retina and visual cortex. In a nominal local sign system, stimulation of each retinal location evokes an identifiable response, but the set of responses is not metrically organized. In an ordinal local sign system, values such as up and down or left and right are specified, and in an interval system, distances between objects may be specified. Quantitative judgments about the oculocentric location of an isolated object require a ratio local sign system, that is, one in which there is a built-in reference point and fiducial line, such as the fovea and the normally vertical meridian.

The headcentric position, orientation or movement of a visual object may be objectively specified in terms of its angle of elevation relative to a transverse plane through the eyes, and its angle of azimuth relative to the median plane of the head. A person making headcentric visual judgments must take account of both oculocentric and eye-in-head information. The bodycentric (torsocentric) position or movement of an object may be objectively specified in terms of the median plane of the head and some arbitrary transverse plane of the body. If no part of the body is in view, bodycentric judgments require the observer to take account of oculocentric

information, eye-in-head information and information from the neck joints and muscles regarding the position of the head on the body. Thus the oculocentric, headcentric and bodycentric reference systems form a hierarchical, or nested, set of egocentric frames as indicated in the second column of table 1. If the body as well as the object being judged is in view, bodycentric judgments are much simpler since they can be done on a purely visual basis without the need to know the positions of the eyes or head. Eye-in-head and head-on-body information provided by afferent or efferent neural signals can, at least in theory, provide nominal, ordinal, interval, or ratio metrics.

Finally, the exocentric position, orientation, or movement of an object is specified with respect to arbitrary coordinates external to the body. Exocentric judgments about an isolated visual object require the observer to take account of oculocentric, eye-in-head and head-on-body information and, in addition, information regarding the position or movement of the body with respect to an external frame. This may involve associating the position of a seen object with, for instance, the position of the noise that it is making. This is a multisensory task. In other cases it may involve relating the position of an object detected by one sense organ with the position of another object detected by a second sense organ. This is an intersensory task (see Howard, 1982, Chapter 11, for more details on this distinction). The vestibular system is the only sense organ that provides direct information about the attitude and movement of the body in inertial space. The otolith organs respond to the static and dynamic pitch and roll of the head with respect to gravity; they provide no information about rotation or position of the head around the vertical axis. The otolith organs also respond to linear acceleration of the body along each of three orthogonal axes, but cannot distinguish between head tilt and linear acceleration. The semi-circular canals provide information about body rotation in inertial space about each of three orthogonal axes. But if rotation is continued at a constant angular velocity, the input from the canals soon ceases. The integral of the motion signal from the canals can provide information about the position of the body, but only with respect to a remembered initial position. If there are two point-objects in view at the same time, exocentric judgments of the distance between them and their relative motion are possible using only oculocentric information. At least three point-objects are required for exocentric visual judgments of direction or orientation based solely on oculocentric information.

In what follows I shall discuss the extent to which perceptual judgments within egocentric and exocentric frames of reference are subject to illusory disturbances and long-term modifications. I shall argue that well-known spatial illusions, such as the oculogyral illusion and induced visual motion have usually been discussed without proper attention being paid to the frame of reference within which they occur, and that this has led to the construction of inadequate theories and inappropriate procedures for testing them.

## 2. THE OCULOCENTRIC FRAME

Any misperception of the oculocentric position or movement of a visual object can arise only as a result of some disturbance of the retinal local sign system or of the oculocentric motion-detecting system. In a geometrical illusion, lines are apparently distorted or displaced when seen in the context of a larger pattern. In a figural aftereffect, a visual test object seen in the neighborhood of a previously seen inspection object appears displaced away from the position of the inspection object. Such effects operate only over distances of about one degree of visual

angle, and the apparent displacement rarely exceeds a visual angle of a few minutes of arc (Kohler and Wallach, 1944). We must conclude that the local sign system is relatively immutable. This is not surprising, since the system depends basically on the anatomy of the visual pathways. Several claims have been made that oculocentric distortions of visual space can be induced by pointing with hidden hand to visual targets seen through displacing prisms (Cohen, 1966; Held and Rekosh, 1963). Others have claimed that these effects were artifactual, and we are left with no convincing evidence that oculocentric shifts can be induced in this way. (See Howard, 1982, p, 501 for a more detailed discussion of this subject.)

The movement after effect is a well-known example of what is almost certainly an oculocentric disturbance of the perception of motion. I will not discuss this topic here.

## 3. THE HEADCENTRIC FRAME

A misjudgment of the headcentric direction or motion of a visual object could arise from a misregistration of the position or motion of either the retinal image or the eyes. In this section I shall consider only phenomena due to misregistration of the position or movement of the eyes.

### 3.1 Illusory Shifts of Headcentric Visual Direction

Deviations of the apparent straight ahead due to misregistered eye position are easy to demonstrate. If the eyes are held in an eccentric position, a visual target must be displaced several degrees in the direction of the eccentric gaze to be perceived as straight ahead. When the observer attempts to look straight ahead after holding the eyes off to one side, the gaze is displaced several degrees in the direction of the previous eye deviation. Attempts to point to visual targets with unseen hand are displaced in the opposite direction. The magnitude of these deviations has been shown to depend on the duration of eye deviation and to be a linear function of the eccentricity of gaze (Hill, 1972; Morgan, 1978; Paap and Ebenholtz, 1976). Similar deviations of bodycentric visual direction occur during and after holding the head in an eccentric posture (Howard and Anstis, 1974). It has never been settled whether these effects are due to changes in afference or to changes in efference associated with holding the eyes in a given posture (see Howard, 1982, for a discussion of this issue). Whatever the cause of these effects, it is evident that the headcentric system is more labile than the oculocentric system. This is what one would expect, because headcentric tasks require the neural integration of information from more than one sense organ.

### 3.2 The Oculogyral Illusion

The oculogyral illusion may be defined as the apparent movement of a visual object while the semicircular canals of the vestibular system are being stimulated (Graybiel and Hupp, 1946). The best visual object is a small point of light in otherwise dark surroundings and fixed with respect to the head. When the vestibular organs are stimulated, as for instance by accelerating the body about the mid-body axis, the point of light appears to race in the direction of body rotation. The oculogyral illusion also occurs when the body is stationary, but the vestibular organs signal that it is turning. This happens, for instance, in the 20 or 30 seconds after the body has been

brought to rest after being rotated. It is not surprising that a point of light attached to the body should appear to move in space when the observer feels that the body is rotating. I shall refer to this perceived motion of the light with the body as the exocentric component of the oculogyral illusion. The exocentric component is not very interesting because it is difficult to see how a rotating person could do other than perceive a light which is attached to the body as moving in space. But even casual observation of the oculogyral illusion reveals that the light appears to move with respect to the head in the direction of body acceleration. This headcentric motion of the light is the headcentric component of the oculogyral illusion.

Whiteside, Graybiel and Niven (1965) proposed that the headcentric component of the oculogyral illusion is due to the effects of unregistered efference associated with the vestibulo-ocular response (VOR) The idea is that when the subject fixates the point of light, VOR engendered by body acceleration is inhibited by voluntary innervation. The voluntary innervation is fully registered by the perceptual system, but the VOR efference is not, and this asymmetry in registered efference causes the subject to perceive the eyes as moving in the direction of body rotation. This misperception of the movement of the eyes is interpreted by the subject as a headcentric movement of the fixated light. To support this theory, we need evidence that the efference associated with VOR is not fully registered by the perceptual system responsible for making judgments about the headcentric movement of visual objects.

For frequencies of sinusoidal head rotation up to about 0.5 Hz, the VOR is almost totally inhibited if the attention is directed to a visual object fixed with respect to the head (Benson and Barnes, 1978). The most obvious theory is that VOR suppression by a stationary object is due to cancellation of the VOR by an equal and opposite smooth pursuit generated by the retinal slip signal arising from the stationary light. This cannot be the whole story because Barr, Schulthies and Robinson (1976) reported that the gain of VOR produced by sinusoidal body rotations decreased to about 0.4 when subjects imagined that they were looking at an object rotating with them. It looks as though VOR efference can be at least partially cancelled or switched off even without the aid of visual error signals (McKinley and Peterson, 1985; Melvill Jones, Berthoz and Segal, 1984). Tomlinson and Robinson (1981) were concerned to account for how an imaginary object can inhibit VOR, but for our present purposes, the more important point is that VOR is not totally inhibited. Perhaps an imagined object is not a satisfactory stimulus for revealing the extent of voluntary control over VOR. We wondered whether an afterimage might be a better stimulus because it relieves subjects of the task of imagining an object and only requires them to imagine that it is stationary with respect to the head. We had already found optokinetic nystagmus (OKN) to be totally inhibited by an afterimage, even though it was not inhibited by an imaginary object. The results of all these experiments are reported in Howard, Giaschi and Murasugi (1988).

Subjects in total darkness were subjected to a rotary acceleration of the whole body of $14°/s^2$ to a terminal velocity of $70°/s$, which was maintained for 60 s. In one condition subjects were asked to carry out mental arithmetic. In a second condition they were asked to imagine an object rotating with the body, and in a third condition, an afterimage was impressed on both eyes just before the trial began and the subject was asked to imagine that it was moving with the body. The same set of conditions was repeated, but with lights on, so that the stationary OKN display filled the visual field. Under these conditions both VOR and OKN are evoked at the same time.

In all conditions the velocity of the slow phase of each nystagmic beat was plotted as a function of time from the instant that the body reached its steady-state velocity. For none of the subjects was VOR totally inhibited at any time during any of the trial periods. For the OKN plus VOR condition, subjects could initially inhibit the nystagmus only partially, even though they could see a moving display, but they could totally inhibit the response after about 30 s, when the VOR signal had subsided.

We propose that VOR is not completely inhibited by an afterimage seen in the dark because the mechanism used to assess the headcentric motion of visual objects does not have full access to efference associated with VOR. Thus the system has no way of knowing when the eyes are stationary. The component of the VOR which cannot be inhibited by attending to an afterimage gives an estimate of the extent to which VOR efference is unregistered by the system responsible for generating voluntary eye movements and for giving rise to the headcentric component of the oculogyral illusion.

## 4. THE EXOCENTRIC FRAME

### 4.1 Vection

Vection is an illusion of self-motion induced by looking at a large moving display and is the clearest example of an exocentric illusion. For instance, illusory self-rotation, or circular-vection, is induced when an upright subject observes the inside of a large vertical cylinder rotating about the mid-body axis (yaw axis). For much of the time the cylinder seems to be stationary in exocentric space and the body feels as if it is moving in a direction opposite to that of the visual display. Similar illusions of self-motion may be induced by visual displays rotating about the visual axis (roll axis) or about an axis passing through the two ears (pitch axis) (Dichgans and Brandt, 1978). Rotation of a natural scene with respect to the head is normally due to head rotation, and the vestibular system is an unreliable indicator of self-rotation except during and just after acceleration. Therefore it is not surprising that scene rotation is interpreted as self-rotation, even when the body is not rotating. There is a conjunction of visual and vestibular inputs into the vestibular nuclei (Waespe and Henn, 1978) and the parietal cortex (Fredrickson and Schwarz, 1977), which probably explains why visual inputs can so closely mimic the effects of vestibular inputs.

4.1.1 *Vection for different postures and axes of rotation* – If the vection axis is vertical, the sensation of self-rotation is continuous and is usually at the full velocity of the stimulus motion. If the vection axis is horizontal, the illusory motion of the body is restrained by the absence of utricular inputs that would arise if the body were actually rotating. Under these circumstances a weakened but still continuous sensation of body rotation is accompanied by a paradoxical sensation that the body has tilted only through a certain angle (Held, Dichgans and Bauer, 1975). Howard, Cheung and Landolt (1987) suspended a subject in various postures within a large sphere that could be rotated about a vertical or horizontal axis and measured the magnitude of vection and illusory body tilt for yaw, pitch and roll vection for both vertical and horizontal orientations of each axis (fig. 1).

For body rotation about both vertical and horizontal axes, yaw vection was stronger than pitch vection, which was stronger than roll vection. When the vection axis was vertical,

sensations of body motion were continuous and usually at, or close to, the full velocity of the rotating visual field. When the vection axis was horizontal, the sensations of body motion were still continuous, but were reduced in magnitude. Also, for vection about horizontal axes, sensations of continuous body motion were accompanied by sensations of illusory yaw, roll, or pitch of the body away from the vertical posture. The mean body tilt was over $20°$, but the body was often reported to have tilted by as much as $90°$. Two subjects in a second experiment reported sensations of having rotated full circle. Held, Dichgans and Bauer (1975) reported a mean illusory body tilt of $14°$. We obtained larger degrees of body tilt, probably because our display filled the entire visual field and because subjects were primed to expect that their bodies might really tilt. In most subjects, illusory backwards tilt produced by pitch vection about a horizontal axis was much stronger than illusory forward tilt . Only two of our 16 subjects showed the opposite asymmetry; that was also reported by Young, Oman and Dichgans (1975).

4.1.2 *Vection and the relative distances of competing displays* – The more distant parts of a natural scene are less likely to rotate with a person than are nearer parts of a scene, so that the headcentric motion of more distant parts provides a more reliable indicator of self-rotation than does motion of nearer objects. It follows that circularvection should be related to the motion of the more distant of two superimposed displays. In line with this expectation Brandt, Wist, and Dichgans (1975) found that vection was not affected by a stationary object in front of the moving display, but was reduced when the object was seen beyond the display. Depth was created by binocular disparity in this experiment, and there is some doubt whether depth was the crucial factor as opposed to the perceived foreground-background relationships of the competing stimuli. Furthermore, the two elements of the display differed in size as well as distance.

Ohmi, Howard and Landolt (1987) conducted an experiment using a background cylindrical display of randomly placed dots which rotated around the subject, and a similar stationary display mounted on a transparent cylinder which could be set at various distances between the subject and the moving display. The absence of binocular cues to depth allowed the perceived depth order of the two displays to reverse spontaneously, even when they were well separated in depth. Subjects were asked to focus alternately on the near display and the far display while reporting the onset or offset of vection. They were also asked to report any apparent reversal of the depth order of the two displays, which was easy to notice because of a slight difference in appearance of the two displays.

In all cases vection was experienced whenever the display that was perceived as the more distant was moving and was never experienced whenever the display perceived as more distant was stationary. Thus circular vection is totally under the control of whichever of two similar displays is perceived as background. This dominance of the background display does not depend on depth cues, because circularvection is dominated by a display that appears more distant, even when it is nearer. We think that perceived distance is not the crucial property of that part of the scene interpreted as background. When subjects focused on the moving display, optokinetic pursuit movements of the eyes occurred, and when they focused on the stationary display, the eyes were stationary. But such a change in the plane of focus had no effect on whether or not vection was experienced, as long as the apparent depth order of the two displays did not change.

Thus sensations of self rotation are induced by those motion signals that are most reliably associated with actual body rotation—namely, signals arising from that part of the scene perceived as background. Vection sensations are not tied to depth cues, which makes sense because depth cues can be ambiguous. Nor are vection sensations tied to whether the eyes pursue one

part of the scene or another, which also makes sense because it is headcentric visual motion that indicates self-motion, which is just as well detected by retinal image motion as by motion of the eyes.

4.1.3 *Circularvection and the central-peripheral and near-far placement of stimuli* – It has been reported that circularvection is much more effectively induced by a moving scene confined to the peripheral retina than by one confined to the central retina (Brandt, Dichgans and Koenig, 1973). In these studies, the central retina was occluded by a dark disc which may have predisposed subjects to see the peripheral display as background, and it may have been this, rather than its peripheral position, which caused it to induce strong vection. Similarly, when the stimulus was confined to the central retina, subjects may have been predisposed to see it as a figure against a ground, which may have accounted for the small amount of vection evoked by it.

Howard et al. (1987) conducted an experiment to test this idea. The apparatus is depicted in figure 2. The subject sat at the center of a vertical cylinder covered with randomly arranged black opaque dots. A 28° square display of dots above the subject's head was reflected by a sheet of transparent plastic onto a matching black occluder in the center of the large display. The central display could be moved so that it appeared to be suspended in front of, in the same plane as, or beyond the peripheral display. In the latter position it appeared as if seen through a square hole. In some conditions, one of the displays moved from right to left or from left to right at 25°/s while the other was occluded. In other conditions both displays were visible, but only one moved and in still other conditions, both displays moved, either in the same direction or in opposite directions. In each condition subjects looked at the center of the display and rated the direction and strength of circularvection.

The results are shown in figure 3. They reveal that, all things being equal, vection is driven better by peripheral stimuli than by a 28° central stimulus  Indeed, it is driven just as well by a moving peripheral display with the center black or visible and stationary as it is by a full-field display. However, if the center of the display is moving in a direction opposite to that of the peripheral part, then vection is reduced. Thus a moving central display can weaken the effect of a moving peripheral display, but not to the extent of reversing vection. If the peripheral part of the display is visible but stationary, then the direction of vection is determined by the central part of the display, but only if the moving central field is farther away than the surround. This result is understandable when we realize that this sort of stimulation is produced, for example, when an observer looks out of the window of a moving vehicle. The moving field seen through the window indicates that the viewer is carried along with the part of the scene surrounding the window on the inside. When the surround is black, vection is still controlled by the movement of the central display, even when it is coplanar with or in front of the surround. The reason for this is probably that a central display in front of a black surround provided virtually no cues to its location in depth and subjects perceived it as being beyond the surrounding black display.

## 4.2 Induced Visual Motion

Induced visual motion occurs when one observes a small stationary object against a larger moving background and was first described in detail by Duncker (1929). For instance, the moon appears to move when seen through moving clouds. There is a form of induced motion in which the stationary object is seen against a frame which moves across it. In this stimulus configuration, the moving frame becomes increasingly eccentric and this may be responsible for some of

the illusory motion of the stationary object. I do not wish to consider the asymmetry effect, so the stimulus I shall consider is one in which the stationary object is seen against a large moving background that either fills the visual field or remains within the confines of a stationary boundary.

Induced visual motion could occur within the oculocentric, the headcentric or the exocentric system. As an oculocentric effect, it could be due to contrast between oculocentric motion detectors. I shall argue that this is not a major cause of the illusion.

As a headcentric effect, induced visual motion could be due to OKN induced by inhibition of the moving background by voluntary fixation on the stationary object. If the efference associated with OKN were not available to the perceptual system, but the efference associated with voluntary fixation were, this should create an illusion of movement in a direction opposite to that of the background motion. This explanation, which I proposed in 1982, is analogous to that proposed by Whiteside, Graybiel and Niven (1965) to account for the oculogyral illusion. It has been championed more recently by Post and Leibowitz (1985) and Post (1986). I believe that the evidence reviewed below shows that this is not the main cause of induced visual motion.

Induced visual motion could be an exocentric illusion. It has been explained that inspection of a large moving background induces an illusion of self-motion accompanied by an impression that the background is not moving. A small object fixed with respect to the observer should appear to move with the observer and therefore to move with respect to the exocentric frame provided by the perceptually stationary background. This possibility was mentioned by Duncker and is, I suggest, the major cause of induced visual motion. I shall now review evidence in favour of this explanation of induced visual motion.

4.2.1 *Inhibition of OKN is neither necessary nor sufficient for induced motion* – In the experiment on circularvection described in section 4.1.2, Ohmi, Howard, and Landolt (1987) showed that vection occurred whenever the more distant of two displays was moving, but never when the more distant display was stationary. When the more distant display moved, vection occurred both when the subjects converged on the moving display and had OKN, and when they converged on the stationary nearer display and inhibited OKN. The important point in the present context is that the nearer stationary display appeared to move with the subject (exocentrically) whenever there was vection, but appeared perfectly stationary when there was no vection. Thus, induced visual motion came and went with vection and did not depend on whether or not OKN was inhibited. McConkie and Farber (1979) reported that a visual display perceived as background induced visual motion in an otherwise similar display perceived as foreground, although they did not relate this to changes in vection.

The theory that ascribes induced visual motion to contrast between oculocentric motion detectors cannot account for these results, because the same relative motion was present when the far display moved and the near display did not, as when the near display moved and the far one did not. According to the oculocentric theory there should have been induced motion in both cases rather than only in the first.

The headcentric theory of induced visual motion that explains the effect in terms of inhibition of involuntary OKN by voluntary efference cannot account for these results either, because induced motion occurred whether or not OKN was inhibited. Furthermore, when a stationary display was seen as the background to a moving display, vection did not occur, even

when subjects attended to the stationary display and inhibited OKN. Thus, whether or not OKN was inhibited had no bearing on whether induced visual motion occurred under these circumstances.

Vection is an exocentric phenomenon, and induced visual motion of stationary elements of the visual display comes and goes with saturated vection. The stationary elements simply look as if they are rotating with the body, not slower and not faster. If vection is fully saturated, the moving scene appears stationary and the body and stationary elements of the scene appear to move exocentrically at the full velocity of the inducing field. Under these circumstances induced visual motion is complete. For instance, if a large scene rotates at $60°/s$, induced visual motion of a stationary object is also that velocity. All this suggests that induced visual motion can be an exocentric effect coupled to vection. Headcentric induced motion may occur in other conditions.

The exocentric theory of induced visual motion nicely explains why there is no loss of accuracy in pointing with unseen hand to a visual target subjected to induced visual motion (Bacon, Gordon and Schulman, 1982; Bridgeman, Kirsch and Sperling, 1981). A headcentric theory of induced motion predicts that pointing would deviate, since any misperception of gaze should be reflected in the bodycentric task of pointing. On the exocentric theory, there should be no loss in pointing accuracy, since pointing is a bodycentric task.

It might be objected that when a single stationary object is placed against a small moving display it exhibits induced motion, although there is no discernable illusion of self-motion. I think this is because the visual consequences of vestibular stimulation have a lower threshold than the sensations of body motion. For instance, it is well known that the oculogyral illusion induced by actual body rotation gives a more sensitive measure of vestibular thresholds than do sensations of body motion (Miller and Graybiel, 1975). When the inducing field is small, induced visual motion is only a fraction of the velocity of the inducing field, but as the size of the inducing field is increased, vection becomes evident and induced visual motion more pronounced until, when the field is sufficiently large, both vection and induced visual motion attain the full value of the velocity of the moving field. When vection and induced visual motion are saturated, the objectively stationary object appears to move in exocentric space at the same velocity as the body, neither getting ahead nor lagging behind. In other words, with large inducing fields there is no perceptible headcentric component of induced visual motion. The stationary object may appear to be headcentrically displaced in the direction of motion of the background, but that is a displacement effect, not an illusory motion. This effect may be related to the well-known fact that, in the absence of a fixation point, the eyes deviate in the direction of the fast phases of OKN (Brecher, et al., 1972; Heckmann and Post, 1986). It is possible that when a visual display is accelerating, the increasing deviation of gaze induces an apparent motion in a stationary object. However, I am dealing here only with illusory visual motion induced by visual displays moving at constant velocity.

4.2.2 *Evidence that OKN efference is perceptually registered* – The fact that a headcentric component of induced visual motion may be absent suggests that efference associated with OKN is available to the perceptual system, unlike that associated with VOR. We recently produced evidence that this is so (Howard, Giaschi and Murasugi, 1988).

Optokinetic nystagmus is induced when a person looks at a moving textured surface. The response cannot be inhibited by voluntary effort, as long as the eyes remain converged on the moving display (Howard and Gonzalez, 1987). However, the response is totally inhibited

if attention is directed to a stationary object superimposed on the center of the display (Murasugi, Howard, and Ohmi, 1986). If the attention is directed to an afterimage imposed on the fovea, OKN may be totally inhibited (Viefhues, 1958; Murasugi, Howard and Ohmi, 1984: Wyatt and Pola, 1984). If the afterimage is regarded as fixed in space, then OKN is inhibited and the after image appears stationary. If the afterimage is regarded as moving with the moving display, then OKN is fully restored. It is easy to understand how a real stationary object allows a person to inhibit OKN; any movement of the eyes with respect to the stationary object generates both a misfoveation (position) signal and a retinal slip (velocity) signal. However, these error signals are not provided by an afterimage, so that some other error signal or an open-loop signal must be used in this case. The effect cannot be due to occlusion of the moving display by the afterimage because OKN was only partially reduced when the center of the display was occluded by a black horizontal band. The more OKN is inhibited, the more the eyes lag behind the moving display and the greater is the relative motion between afterimage and display. However, although relative motion is minimum when OKN gain is one, it has no maximum value because it would continue to increase if the eyes were to move in a direction opposite to that of the display. In other words, the degree of relative motion between afterimage and moving display does not indicate when the eye velocity is zero. A partial loss of gain of OKN found in some subjects when imagining a head-fixed object is presumably due to the injection of a voluntary command into the eye movement signal. But this effect accounts for only a small part of the complete suppression of OKN by an afterimage.

The inhibition of OKN by an afterimage could be due to the production of a voluntary efferent command of opposite sign which cancels the OKN efference signal. If the voluntary mechanism had only partial access to the efference controlling OKN, then it would not be able to produce a matching command and bring the eyes to a stop and at the same time perceive the afterimage as stationary with respect to the head. An object imagined in the plane of the display is ineffective, and this must be because it provides no confirming impression of a stationary object once OKN efference has been cancelled. In the absence of such an object, there is an overriding necessity to stabilize the image of the moving stimulus.

4.2.3 *Induced visual motion in several directions simultaneously* – Visual motion has been reported to be induced by stimuli moving simultaneously in two directions. For instance, Nakayama and Tyler (1978) reported that a pair of parallel lines pulsing in and out in opposite directions induced an apparent pulsation of a pair of stationary lines placed between them. However, the apparent velocity of this induced motion was only about $0.1°$/s and the effect may have been an oculocentric effect akin to the figural aftereffects. But in any case, the exocentric theory of induced visual motion can account for induced visual motion in more than one direction. For instance, an outwardly expanding textured surface induces forward linear vection (Anderson and Braunstein, 1985). Ohmi and Howard (1988) found that forward linear vection induced by a looming display, and the accompanying induced visual motion of a superimposed stationary display occurred only if the looming display appeared more distant than the stationary display. According to the oculocentric theory of induced visual motion, the depth order of the two displays should not matter. A theory of induced visual motion based on the inhibition OKN cannot account for induced visual motion produced by looming displays, since such displays do not invoke OKN.

It is possible that there is a headcentric component to induced visual motion under certain circumstances, such as when a visual display is accelerating or becoming more eccentric. But the

above evidence strongly suggests that the major part of induced visual motion induced by large moving fields under steady conditions is exocentric and is a simple consequence of vection. Visual motion induced under these circumstances can be 100% of the velocity of the inducing field. Furthermore, visual motion may be induced in a stationary display that fills the visual field if the display is perceived as a foreground in front of a large moving background.

# REFERENCES

Anderson, G. J. and Braunstein, M. L. (1985) Induced self-motion in central vision. J. Exp. Psychol., Hum. Percept. Perform., 11, 122-132.

Bacon, J. H., Gordon, A. and Schulman, P. H. (1982) The effect of two types of induced-motion displays on perceived location of the induced target. Percept. Psychophys., 32, 353-359.

Barr, C. C., Schulthies, L. W. and Robinson, D.A. (1976) Voluntary, nonvisual control of the human vestibulo-ocular reflex. Acta Otolaryngol., 81, 365-375.

Benson, A. J. and Barnes, G. R. (1978) Vision during angular oscillation: The dynamic inter-action of visual and vestibular mechanism. Aviat. Space Environ. Med., 49, 340-345.

Brandt, T., Dichgans, T. and Koenig, E. (1973) Differential effects of central versus peripheral vision on egocentric and exocentric motion perception. Exp. Brain Res., 16, 476-491.

Brandt, T., Wist, E. R. and Dichgans, J. (1975) Foreground and background in dynamic spatial orientation. Percept. Psychophys., 17, 497-503.

Brecher, G. A., Brecher, M. H., Kommerell, G., Sauter, F. A. and Sellerbeck, J. (1972) Relation of optical and labyrinthean orientation. Optica Acta, 19, 467-471.

Bridgeman, B., Kirsch, M. and Sperling, A. (1981) Segregation of cognitive and motor aspects of visual function using induced motion. Percept. Psychophys., 29, 336-342.

Cohen, H. B. (1966) Some critical factors in prism adaptation. Amer. J. Psychol., 79, 285-290.

Dichgans, J. and Brandt, T. (1978) Visual-vestibular interaction: Effects on self-motion perception and postural control. In R. Held, H. W. Leibowitz and H. L. Teuber (eds). Handbook of Sensory Physiology, Vol. VIII. New York; Springer. pp 755-804.

Duncker, K. (1929) Uber induzierte Bewegung. Psychol Forsch., 22, 180-259.

Fredrickson, J. M. and Schwarz, D. W. F. (1977) Vestibulo-cortical projection. In R. F. Naunton (Ed.) The Vestibular System. New York: Academic Press, pp 203-210.

Graybiel, A. and Hupp D. I. (1946) The oculo-gyral illusion. J. Aviat. Med., 17, 3-27.

Heckmann, T. and Post, R. B. (1986) Induced motion and apparent straight ahead during pro-longed stimulation. Invest. Ophthal. Vis Sci., 27, (ARVO Suppl.), 346.

Held, R., Dichgans, J. and Bauer, J. (1975) Characteristics of moving visual areas influencing spatial orientation. Vision Res., 15, 357-365.

Held, R. and Rekosh, J. (1963) Motor-sensory feedback and geometry of visual space. Science, 141, 722-723.

Hill, A. L. (1972) Direction constancy. Percept. Psychophys., 11, 175-178.

Howard, I. P. (1982) Human Visual Orientation. Chichester; Wiley. pp 388-397.

Howard, I. P. and Anstis, T. (1974) Muscular and joint-receptor components in postural persistence. J. Exp. Psychol., 103, 167-170.

Howard, I. P., Cheung, B. and Landolt, J. (1987) Influence of vection axis and body posture on visually-induced self rotation and tilt. Proceedings of the AGARD Conference on Motion Cues in Flight Simulation and Simulator Sickness. Brussels.

Howard, I. P., Giaschi, D. and Murasugi, C. M. (1988) Inhibition of optokinetic and vestibular nystagmus by head-fixed stimuli. Exp. Brain Res., submitted for publication.

Howard, I. P. and Gonzalez, E. G. (1987) Optokinetic nystagmus in response to moving binocularly disparate stimuli. Vision Res., 27, 1807-1817.

Howard, I. P., Ohmi, M., Simpson, W. and Landolt, J. (1987) Vection and the spatial disposition of competing moving displays. Proceedings of the AGARD Conference on Motion Cues in Flight Simulation and Simulator Sickness. Brussels.

Kohler, W. and Wallach. H. (1944) Figural aftereffects: An investigation of visual processes. Proc. Amer. Philos. Soc., 88, 269-357.

McConkie, A. B. and Farber, J. M. (1979) Relation between perceived depth and perceived motion in uniform flow fields. J. Exp. Psychol. Hum. Percept. Perform., 5, 501-508.

McKinley, P.A. and Peterson, B. W. (1985) Voluntary modulation of the vestibuloocular reflex in humans and its relation to smooth pursuit. Exp. Brain Res., 60, 454-464.

Melvill Jones, G., Berthoz, A. and Segal, B. (1984) Adaptive modification of the vestibuloocular reflex by mental effort in darkness. Brain Res., 56, 149-153.

Miller, E. F. and Graybiel, A. (1975) Thresholds for the perception of angular acceleration as indicated by the oculogyral illusion. Percept. Psychophys., 17, 329-332.

Morgan, C. L. (1978) Constancy of egocentric visual direction. Percept. Psychophys., 23, 61-68.

Murasugi, C. M., Howard, I. P. and Ohmi, M. (1984) Inhibition of optokinetic nystagmus by an afterimage. Invest. Ophthal. Vis. Sci., 25 (ARVO Suppl), 185.

Murasugi, C. M., Howard, I. P. and Ohmi, M. (1986) Optokinetic nystagmus: The effects of stationary edges, alone and in combination with central occlusion. Vision Res., 26, 1155-1162.

Nakayama, K. and Tyler, C.W. (1978) Relative motion induced between stationary lines. Vision Res., 18, 1663-11668.

Ohmi, M., Howard, I. P. and Landolt, J. P. (1987) Circular vection as a function of foreground-background relationships. Perception, 16, 17-22.

Ohmi, M. and Howard, I. P. (1988) Effect of stationary objects on illusory forward self-motion induced by a looming display. Perception, submitted for publication.

Paap, K. R. and Ebenholtz, S. M. (1976) Perceptual consequencies of potentiation in the extraocular muscles: An alternative explanation for adaptation to wedge prisms. J. Exp. Psychol., 2, 457-468.

Post, R. B. (1986) Induced motion considered as a visually induced oculogyral illusion. Perception, 15, 131-138.

Post, R. B. and Leibowitz, H. W. (1985) A revised analysis of the role of efference in motion perception. Perception, 14, 631-643.

Tomlinson, R. D. and Robinson, D.A. (1981) Is the vestibulo-ocular reflex cancelled by smooth pursuit? In Fuchs, A. and Becker, J. (Eds.) Progress in Oculomotor Research. Elsevier: North Holland.

Viefhues, T. K. (1958) Uber das Fixieren eines fovealen Nachbildes. von Graefes Archiv. fur Ophthal., 160, 161-163.

Waespe, W. and Henn, V. (1978) Conflicting visual-vestibular stimulation and vestibular nucleus activity in alert monkeys. Exp. Brain Res, 33, 203-211.

Whiteside, T. C. D., Graybiel, A. and Niven, J. I. (1965) Visual illusions of movement. Brain, 88, 193-210.

Wyatt, H. J. and Pola, J. (1984) Suppression of optokinesis by a stabilized target occurs without "looking at" the target. Invest. Ophthal. Vis. Sci., 25 (ARVO Suppl.), 230.

Young, L. R., Oman, C. M. and Dichgans. J. M. (1975) Influence of head orientation on visually induced pitch and roll sensations. Aviat. Space Environ. Med., 46, 246-269.

## TABLE 1.– FRAMES OF REFERENCE FOR VISUAL SPATIAL JUDGMENTS. RF IS SHORT FOR REFERENCE FRAME AND O IS SHORT FOR STIMULUS OBJECT

| TYPE | SENSORY COMPONENTS | EXAMPLES |
|---|---|---|
| **EGOCENTRIC** O and RF internal | | |
| PROPRIOCEPTIVE | Sense of position of body parts | Point to the toe |
| **EGOCENTRIC** O external, RF internal | | |
| OCULOCENTRIC | Retinal local sign (plus stereo vision) | Fixate an object, Place a line on a retinal meridian |
| HEADCENTRIC | Eye position + local sign | Place an object in the median plane of the head |
| BODYCENTRIC (Body not in view) | Neck + eye position + local sign | Align a stick to the unseen toe. Place object to left of body |
| BODYCENTRIC (Body in view) | Relative local sign | Align a stick to the seen toe |
| **EXOCENTRIC** O internal, RF external | Sensed body part and external reference | Align the arm with gravity. Point North |
| **EXOCENTRIC** O and RF external | | |
| SINGLE POINT OR LINE | No exocentric judgments possible | |
| VISUAL OBJECTS | Relative local sign | Place object A East of object B. Align three objects |
| MULTISENSORY | One object detected by two senses | Associate the sight and sound of object |
| INTERSENSORY | Visual and non-visual objects compared | Set a line vertical. Point a line to an unseen sound |
| **HETEROCENTRIC** RF internal-external | | |
| GEOGRAPHICAL | Object-to-self plus landmark | Judge that an object is East of the self |
| GRAVITATIONAL | Object-to-self plus gravity | Judge that an object is above the head |
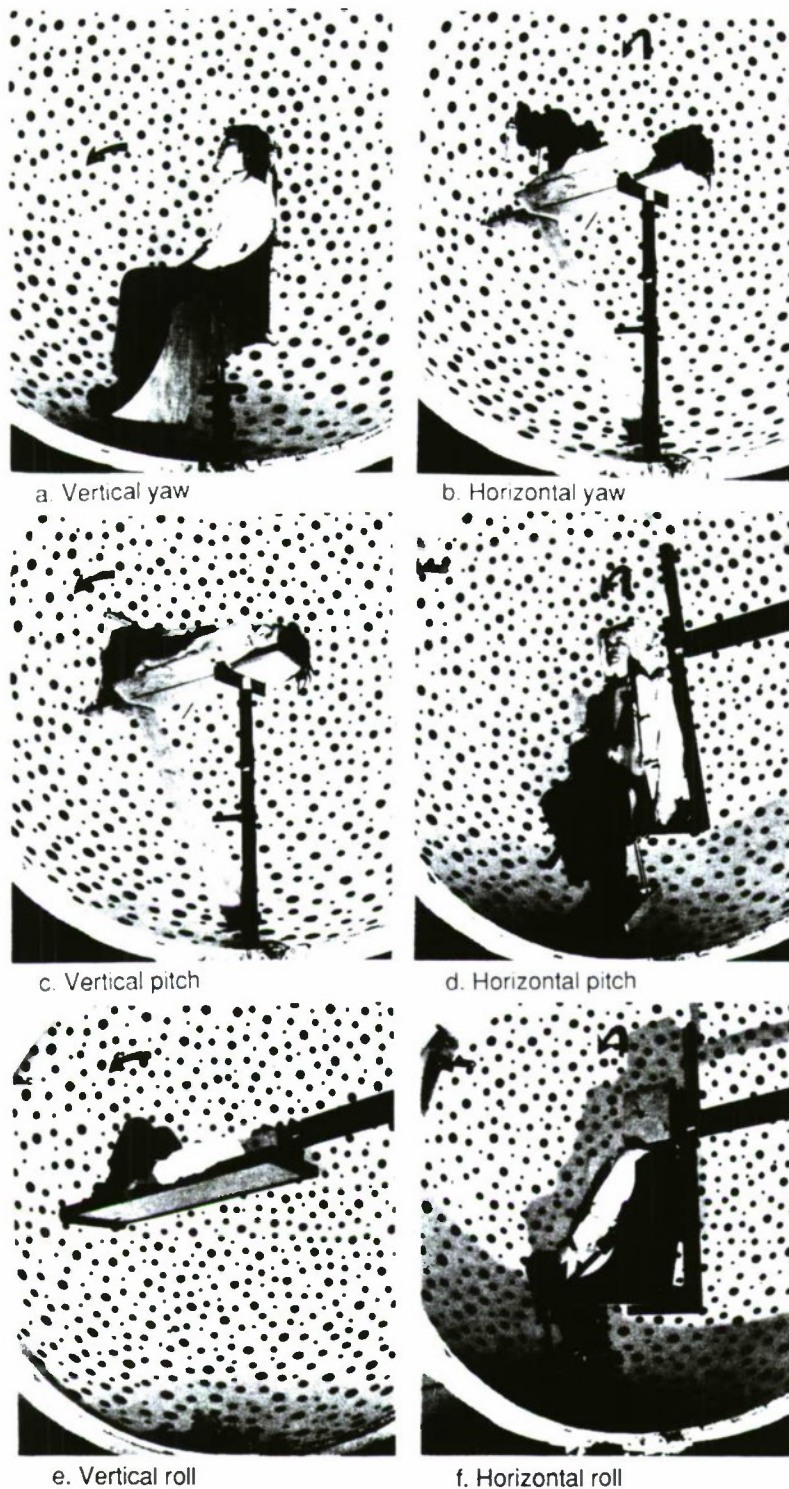
Figure 1.– The set of postures and vection axes use by Howard, Cheung and Landolt (1987) to study vection and illusory body tilt. The subject is seen through the open door of the 3m diameter sphere which could be rotated about either the vertical or horizontal axis. The subject was supported in different postures by air cushions and straps (not shown) so as to produce the six possible combinations of vection axis (yaw, pitch and roll) and gravitational orientation of the axis.
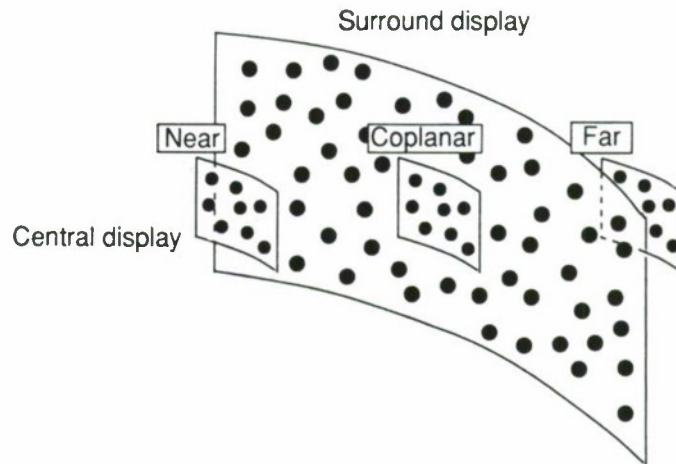
Figure 2.– A diagrammatic representation of the displays use by Howard, Simpson and Landolt (1987) to study the interaction between central-peripheral and far-near placement of two displays in generating circularvection. The two displays could be moved in the same or in opposite directions, or one of them could be stationary or blacked out.
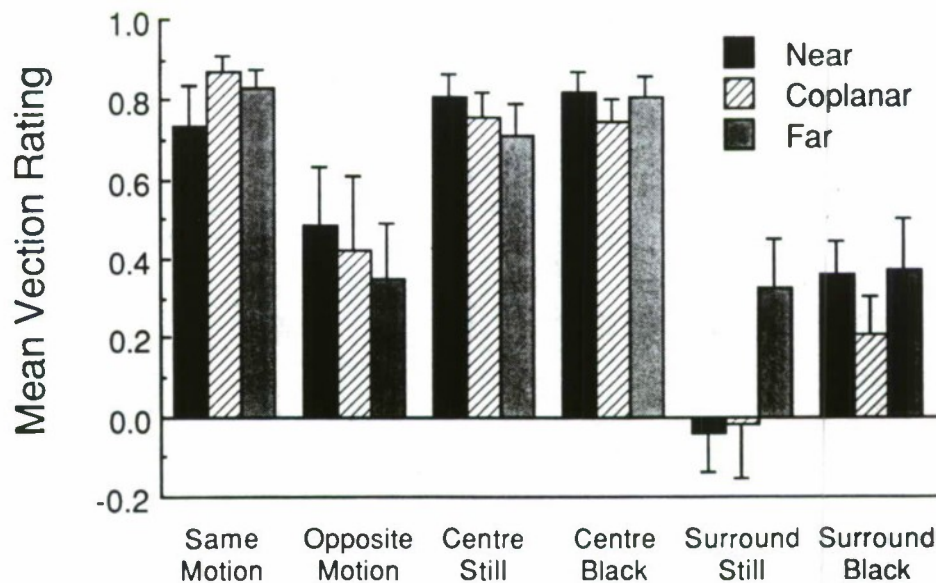


Figure 3.– Mean vection ratings of nine subjects plotted as a function of the relative depth between the central and peripheral parts of the display and the type of display. A vection rating of 1.0 signifies full vection in a direction opposite to the motion of the display. When the two parts of the display moved in opposite directions, the motion of the peripheral part was taken a reference. The error bars are standard errors of the mean.

# COMMENTS ON TALK BY IAN HOWARD

Thomas Heckmann
Human Performance Laboratory
Institute for Space and Terrestrial Science
York University, North York, Ontario, Canada M3J 1P3

Robert B. Post
Department of Psychology
University of California, Davis, CA 95616

Induced visual motion is the name assigned a group of phenomena which can be described with more or less the same words: "illusory motion of stationary contours opposite the direction of moving ones." As Dr. Howard has pointed out, it is possible that oculocentric, headcentric and exocentric mechanisms generate experiences which may be described by the words "induced visual motion." We have found Dr. Howard's framework very helpful in organizing our thoughts about the multiple sources of these apparently similar phenomena. We also accept that some forms of induced visual motion may depend on vection and cannot be explained by suppression of nystagmus (e.g., phenomenal tilt of a stationary stimulus during roll vection induced by a contoured disc rotating in a frontal plane). We are less certain than Dr. Howard, however, that there is only one mechanism for induced visual motion.

In Dr. Howard's study, phenomenal motion of a stationary display which was positioned in front of a moving display occurred only when there was vection. We have reliably obtained induced visual motion of small fixation targets in the complete absence of vection (Post and Heckmann, 1987; Post and Chaderjian, 1988; Heckmann and Post, 1988). Dr. Howard would likely explain this finding with his statement that "...visual consequences of vestibular stimulation have a lower threshold than sensations of bodily motion." We agree wholeheartedly: optokinetic afternystagmus (OKAN), which is a good indicator of the vestibular effects of visual stimulation, has been found at moving-contour velocities too low to elicit vection (Koenig, Dichgans and Schmucker, 1982). We have also reliably obtained OKAN after exposure to a moving-contour stimulus which elicits no vection (Heckmann and Post, 1988). In fact, induced visual motion may be elicited by a single moving dot stimulus (Post and Chaderjian, 1988) which is not capable of producing vection.

If induced visual motion occurs because a perceptually registered voluntary signal for fixation opposes an unregistered involuntary signal for optokinetic nystagmus, then the illusion should reflect known dynamic properties of the optokinetic system. That is, the magnitude of induced visual motion will be proportional to the nystagmus signal being opposed. Induced visual motion should therefore vary across stimulation in the same way that nystagmus varies, but have the opposite directional sign. Our efforts to disconfirm this prediction have so far failed. Induced visual motion is correlated with OKAN of opposite directional sign across variations in stimulus illuminance and velocity (Post, 1986). The magnitude of induced visual motion increases along with the slow-phase velocity of OKAN with increasing stimulus duration. The illusion also decays and reverses direction along with OKAN after stimulus termination. Further, both responses show an increased tendency to reverse direction following stimulation in the presence of a fixation target rather than after stimulation without fixation (Heckmann and Post, 1988).

Induced visual motion is not the only motion illusion involving visual fixation of moving or stationary targets which can potentially be explained by interaction of voluntary and involuntary eye-movement signals. These illusions include autokinesis, the Aubert-Fleischel effect, the Filehne Illusion, and several others (Post and Leibowitz, 1985). Induced visual motion, however, provides a particularly good model for testing the eye-movement hypothesis, since a good deal is known about the dynamics of visually induced involuntary eye movements. We have not been so much interested in "championing" a particular explanation of induced visual motion, therefore, as we have been to test the existence and applicability of a particular mechanism. Of course, since we are using a well-known illusion as our model, we must also explore the applicability of alternative explanations of induced visual motion to our results.

With further reference to the origin of induced visual motion in vection, therefore, we recently reported a dissociation between the two illusions (Post and Heckmann, 1987). Briefly, fixation of a target located 10° left of the midline during exposure to rightward-moving background contours reliably increased the magnitude of induced visual motion. This finding is consistent with the idea that extra voluntary efference is needed to maintain a leftward as compared to a straight-ahead gaze during rightward motion of background contours. Vection, however, was reduced when a fixation target was made available, and further reduced when the target was placed 10° left of the midline. We emphasize that this dissociation does not reject the idea that some form of induced visual motion originates with vection, only the idea that all of induced visual motion originates with vection.

# REFERENCES

Heckmann, T., and Post, R. B. (1988) Induced motion and optokinetic afternystagmus: parallel response dynamics with prolonged stimulation. Vision Res., in press.

Koenig, E., Dichgans, J., and Schmucker, D. (1982) The influence of circularvection (CV) on optokinetic nystagmus (OKN) and optokinetic afternystagmus (OKAN). In Functional Basis of Oculomotor Disorders (ed. by G. Lennerstrand, D. S. Zee, and E. L. Keller), Pergamon, Oxford.

Post, R. B. (1986) Induced motion considered as a visually-induced oculogyral illusion. Perception, 15, 131-138.

Post, R. B., and Chaderjian, M. (1988) The sum of induced and real motion is not a straight path. Percept. Psychophys., 43, 121-124.

Post, R. B., and Heckmann, T. (1987) Experimental dissociation of vection from induced motion and displacement of the apparent straight-ahead. Invest. Ophthal., 28, Supp., 311.

Post, R. B., and Leibowitz, H. W. (1985) A revised analysis of the role of efference in motion perception. Perception, 15, 131-138.

# DISTORTIONS IN MEMORY FOR VISUAL DISPLAYS

Barbara Tversky
Stanford University
Stanford, California

## ABSTRACT

Systematic errors in perception and memory present a challenge to theories of perception and memory and to applied psychologists interested in overcoming them as well. The present paper reviews a number of systematic errors in memory for maps and graphs, and accounts for them by an analysis of the perceptual processing presumed to occur in comprehension of maps and graphs.

Visual stimuli, like verbal stimuli, are organized in comprehension and memory. For visual stimuli, the organization is a consequence of perceptual processing, which is bottom-up or data-driven in its earlier stages, but top-down and affected by conceptual knowledge later on. Segregation of figure from ground is an early process, and figure recognition later; for both, symmetry is a rapidly detected and ecologically valid cue. Once isolated, figures are organized relative to one another and relative to a frame of reference. Both perceptual (e.g., salience) and conceptual factors (e.g., significance) seem likely to affect selection of a reference frame.

Consistent with the analysis, subjects perceived and remembered curves in graphs and rivers in maps as more symmetric than they actually were. Symmetry, useful for detecting and recognizing figures, distorts map and graph figures alike. Top-down processes also seem to operate in that calling attention to the symmetry vs. asymmetry of a slightly asymmetric curve yielded memory errors in the direction of the description. Conceptual frame of reference effects were demonstrated in memory for lines embedded in graphs. In earlier work, the orientation of map figures was distorted in memory toward horizontal or vertical. In recent work, graph lines, but not map lines, were remembered as closer to an imaginary 45° line than they had been. Reference frames are determined by both perceptual and conceptual factors, leading to selection of the canonical axes as a reference frame in maps, but selection of the imaginary 45° line as a reference frame in graphs.

## DISTORTIONS

With the best of intentions, scientists, newspaper editors, and textbook authors select graphic displays to present their ideas more clearly and more vividly to their readers. Nevertheless, some of the effects are not only unintended, but unwanted. For example, in figure 1, presumably the striping on the bars was selected to differentiate the bars, not to instantiate the herringbone illusion, where straight lines are perceived as tilted (this example comes from Schultz, 1961 through Kruskal, 1982). In figure 2 (from the business section of the August 2, 1987, *New York Times*), the graphic artist wanted to contrast two related sets of numbers, the debt and the debt service ratio, year by year. I don't think that the graphic artist intended to create a figure with such a strong tendency to reverse that it makes it difficult to focus on any one section of the graph. Figure 3 takes us from the realm of perceptual illusions to experiments in judgment by Cleveland, Diaconis, and McGill (1982). These statisticians asked knowledgeable subjects to estimate correlations from

scatter plots and found that higher estimates were given when the point cloud was smaller (or the frame larger). Figure 4, popularized by Tufte (1983) and reprinted by Wainer (1980), is taken from the *Washington Post* of October 25, 1978. Here, the graphic artist probably thought it would be clever to represent the metaphor of the diminishing dollar quite literally. However, only the *length* of the dollar represents the decline of purchasing power, not the *area*, yet it is the area that is picked up by the human observer. So, although the Carter dollar purchases a bit less than half of the Eisenhower dollar, the Carter dollar looks less than a quarter of the area of the Eisenhower dollar.

The next example of distorted perception brings me to research in my laboratory. Let me first tell you about a number of different phenomena we have studied, and then I will try to account for them in an analysis of perceptual organization, where both perceptual and conceptual factors are operative. First, I will discuss examples of perceptual factors. Jennifer Freyd and I (1984) asked subjects to look at figures like that at the top of figure 5, and then decide whether it was more similar to a slightly more symmetric figure or to an equally different, but slightly less symmetric, figure. When we selected nearly symmetric figures like that one, subjects nearly always chose the more symmetric alternative as the more similar. What's more, when subjects were asked to select which of the bottom figures was identical to the top figure, subjects were faster to select the identical figure when the alternative figure was *less* symmetric than the original (as in fig. 5) than when it was *more* symmetric than the original. These effects obtained for nearly symmetric figures, but not less symmetric ones. That was rather complicated, but these experiments, and others like them (see Riley, 1962, and Freyd and Tversky, 1984, for reviews) suggest that there is a symmetry bias in perception. Not only do viewers rapidly detect symmetry, but they also perceive nearly symmetric figures as more symmetric than they are. That is, small deviations from symmetry are overlooked. Human faces, for example, are rarely perfectly symmetric, though we think of them as such. The outer men in figure 6 (taken from Neville, 1977, p. 335), for example, are actually the same man at the same time. The two outer pictures were constructed by taking the right and left halves of the actual face in the center, and reproducing them in mirror image. It is only by seeing how different the two constructed symmetric faces are that we become aware of the asymmetry of the original face.

Diane Schiano and I (1987 manuscript, "Distortions memory for graphs and maps") looked for and found distortions toward symmetry in memory for maps and graphs. We presented maps or graphs like those in figure 7 to different groups of subjects. Sometimes, the subjects were asked to sketch the curves of the graphs or the rivers of the maps, and other times, they were asked questions about the content of the maps or graphs. This was done to induce a natural comprehension attitude toward the figures, and to prevent subjects from simply memorizing line shapes. We then asked judges who knew nothing about the hypotheses to rate whether the drawn curves and rivers were more or less symmetric than the original ones. The remembered curves, whether in maps or graphs, were judged more symmetric than the originals. These errors in the direction of symmetry, however, apparently occur in perception, not in memory. We asked another group of subjects to copy the curves, and the copied curves were also judged to be more symmetric than the originals, and to the same degree. The first effect to be accounted for, then, is a tendency to perceive nearly symmetric figures as more symmetric than they actually are.

For the next two effects, I turn to maps. In figure 8 are two maps of the world; which one is correct? If you are like the subjects I have run, most of you will pick the bottom one; that is, the incorrect one. Let me give you another chance. In figure 9 are two maps of the Americas; my apologies to Central America, which was excised not because of the political situation, but for

visual reasons. Again, which map is the correct one? And again, I will predict that most of you will prefer the left, incorrect, one. Why do the incorrect maps look better? Basically, because the incorrect ones are more *aligned*. In the incorrect map of the world, the U.S. and Europe and South America and Africa are more aligned than they are in true map. And in the incorrect map of the Americas, North and South America are more aligned. I found memory errors in the direction of greater alignment for these maps, for directions between major cities on them, for artificial maps, and for visual blobs (Tversky, 1981). Others have found similar results (e.g., Byrne, 1979).

The second prevalent error I have found in maps I termed *rotation*. I asked a group of subjects to place a cut-out of South America in a frame where the canonical directions, north-south and east-west, corresponded, as usual, to the vertical and horizontal sides of the frame (fig. 10). Although the actual orientation is on the right, most of the subjects uprighted South America to the angle of the left-hand figure, or even more so. Not only South America is perceived as tilted. Those of you who live in the Bay Area, or who arrived from the San Francisco airport may think that you drove southwest to Monterey. Most of my local respondents made mistakes like that; for example, thinking that Berkeley is east of Stanford and Santa Cruz is west of Palo Alto. Not so, as this true map of the area shows (fig. 11). Just as for alignment, I have found memory errors of rotation toward the axes for real map figures, for directions between cities on them, for roads, for artificial maps, and for visual blobs (Tversky, 1981). Unlike the symmetry distortion, the distortions produced by alignment and rotation are stronger in memory than in perception; that is, small tendencies toward alignment and rotation appeared in a copy task, but much greater errors appeared in a memory task.

Until now, we have demonstrated that there is a bias toward symmetry in both maps and graphs that appears in perception and is preserved in memory. I have also demonstrated, primarily in maps, biases toward alignment with other figures and rotation to a vertical/horizontal frame of reference that appear slightly in perception and stronger in memory. Now is the time to start to account for these systematic errors by an analysis of perceptual organization, or more specifically, by the effects of perceptual factors in perceptual organization (fig. 12). One of the earliest forms of spatial organization is distinguishing figures from grounds. Because figures are more likely to have symmetry, closure, and other, similar properties than backgrounds, these are valuable cues to figureness (e.g., Hochberg, 1978; Koffka, 1935; Kohler, 1929; Wertheimer, 1958). Symmetry, or near-symmetry, is rapidly and easily detected (e.g., Barlow and Reeves, 1979; Chipman and Mendelson, 1979; Carmody, Nodine, and Locher, 1977; Corballis, 1976). Thus, because of its usefulness in figure discrimination, symmetry seems to be rapidly detected and small deviations from symmetry are overlooked so that nearly symmetric figures are coded and remembered as more symmetric than they really are. Now for anchoring figures in space. In an empty field, figures appear to float, a phenomenon well-known to star-gazers, called the autokinetic effect. In order to perceive and remember the locations of figures, it is useful to anchor them to other figures and/or to a frame of reference. In fact, given that perceivers and the world are rarely static, this seems to be the only way to organize the elements of a scene. Although valuable in locating and orienting figures, anchors pull figures closer to them in memory, yielding systematic errors. Map bodies and graph curves are figures on backgrounds; they are often nearly symmetric, they appear sometimes with other figures, and typically appear in a reference frame. Thus, the analysis of distortion in terms of perceptual organization applies to maps and graphs, and accounts for the errors of symmetry, alignment, and rotation.

This, briefly, is the perceptual analysis. Now, I'd like to present two cases where, we believe, conceptual factors enter into the perceptual analysis of maps and graphs and yield further distor-

tions. This work was also done with Diane Schiano. The first effect brings us back to symmetry. The graph curves we asked subjects to study were slightly, but noticeably, less than symmetric. Given that people perceive such curves as more symmetric than they really are, we wondered if we could weaken or strengthen that belief or perception by an accompanying description of the curve, and consequently alter people's memory of the curve. Again, we presented a variety of graphs for subjects to remember, and tested memory either by asking subjects to draw the graphs or to describe some aspect of the relation depicted by the graph. This time the graphs also included descriptions of the functions. For the nearly-symmetric curve of interest, half the subjects received a description emphasizing its symmetry, that is, "Notice that the curve rises smoothly and falls smoothly." The other subjects received a description emphasizing its asymmetry, that is, "Notice that the curve rises sharply and falls slowly." The curves drawn from memory were given to judges who were unaware of the experimental conditions. The results were just as expected: when attention was directed to the symmetry of the curves, remembered curves were drawn more symmetric than when attention was drawn to the asymmetry of the curve. This result is reminiscent of one of the truly classic experiments in psychology, that of Carmichael, Hogan and Walter (1932).

The second conceptual factor is more subtle, and addresses the issue of what determines the frame of reference. In the absence of any conceptual or meaningful factors, there are often perceptual factors that provide a frame of reference. The typically horizontal and vertical lines of the actual frame of a picture are one example (e.g., Howard and Templeton, 1971). For an environment, the natural vertical plane, up-down, and the two natural horizontal planes, left-right and front-back, form a reference frame; when this is reduced from two to three dimensions, the front-back dimension drops out (e.g., Clark, 1973), usually leaving the horizontal and vertical axes of the picture frame as a reference frame. For maps, there is an additional conceptual factor that is typically perfectly correlated with the perceptually salient axes, namely the canonical directions, north-south and east-west. Thus far, the evidence for alignment has come either from maps and environments, where both perceptual and conceptual factors suggest the horizontal and vertical as a reference frame, or from visual blobs, where perceptual factors suggest the horizontal and vertical.

Schiano and I wondered if simple straight-line functions at various angles in x-y coordinates would be anchored to those coordinates, and thus distorted toward them. Of course, the x-y coordinates form a natural reference frame for graph functions, but unlike streets, graphed functions are rarely perfectly horizontal or vertical. Moreover, there is another reference frame for graphed lines, the (in this case) implicit 45° line. This is the identity line, where x=y, and as such it provides a very important reference point for graphed lines. Above it are steep rises, and below it are shallow ones. The experiments we ran were very similar to the previous graph experiments: there were critical stimuli and distractors, and the memory task was designed to elicit comprehension of content, not just remembering the line. The exact same stimuli were presented as maps to another group of subjects. Subjects were told that the angled lines were paths or short-cuts; they weren't very convincing maps, as can be seen in figure 13. In contrast to the prior work on maps showing alignment to the closest axis, horizontal or vertical, the graph lines were remembered as closer to the imaginary 45° line than they actually were. The map lines showed no systematic distortion, and differed considerably and significantly from the graph lines. We ran this study again, this time using dotted graph lines rather than filled ones. Again, graph lines were remembered as closer to the forty-five degree line, and map lines showed no systematic distortion. This is evidence, we believe, for conceptual factors that influence selection of frame of reference and thereby affect the perceptual analysis, representation, and memory of visual displays.

I have presented a perceptual analysis of figure detection and organization. Both these processes can lead to systematic distortions, which were demonstrated in perception and memory of maps and graphs. Conceptual factors were also shown to affect the perceptual analysis and encoding of visual scenes, and to also yield errors of memory, the description of symmetry in one case, and the selection of a frame of reference in the other. The bottom line is "What you see ISN'T what you get."

# REFERENCES

Barlow, H. B., and Reeves, B. C. (1979). The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. Vision Res., 19, pp. 783-793.

Byrne, R. W. (1979). Memory for urban geography. Quart. J. Exp. Psychol., 31, pp. 147-154.

Carmichael, L. Hogan, H. P., and Walter, A. A. (1932). An experimental study of the effect of language on the reproduction of visually perceived forms. J. Exp. Psychol., 15, pp. 73-86.

Carmody, D. P., Nodine, C. F., and Locher, P. J. (1977). Global detection of symmetry. Percept. Motor Skills, 45, pp. 1267-1273.

Chipman, S. F., and Mendelson, M. J. (1979). Influence of six types of visual structure on complexity judgments in children and adults. J. Exp. Psychol.: Human Percep. Perform., 5, pp. 365-378.

Clark, H. H. (1973). Space, time, semantics, and the child. In T. E. Moore (Ed.), Cognitive development and the acquisition of language. New York: Academic Press.

Cleveland, W. S., Diaconis, P., and McGill, R. (1982). Variables on scatterplots look more highly correlated when the scales are increased. Science, 216, pp. 1138-1141.

Corballis, M. C. (1976). The psychology of left and right. Hillsdale, NJ: Erlbaum.

Freyd, J., and Tversky, B. (1984). Force of symmetry in form perception. Amer. J. Psychol., 97, pp. 109-126.

Hochberg, J. E. (1978). Perception. Second Edition. Englewood Cliffs, NJ: Prentice-Hall.

Howard, I. P., and Templeton, W. B. (1971). Human spatial orientation. London: Wiley.

Koffka, K. (1935). Principles of Gestalt Psychology. New York: Harcourt, Brace.

Kohler, W. (1929). Gestalt Psychology. New York: Liveright.

Kruskal, W. (1982). Criteria for judging statistical graphics. Utilitas Mathematica, 21B, pp. 283-310.

Neville, A. C. (1977). Symmetry and asymmetry problems in animals. In R. Duncan and M. Weston-Smith (Eds.), The encyclopedia of ignorance. New York: Pergamon, pp. 331-338.

Riley, D. A. (1962). Memory for form. In L. Postman (Ed.), Psychology in the making. New York: Knopf, pp. 402-465.

Schultz, G. M. (1961). Beware of diagonal lines in bar graphs. Prof. Geogr., 13, pp. 28-29.

Tufte, E. R. (1983). The visual display of quantitative information. Cheshire, CO: Graphics Press.

Tversky, B. (1981). Distortions in memory for maps. Cognitive Psychol., 13, pp. 407-433.

Wainer, H. (1980). Making newspaper graphs fit to print. In P. A. Kolers, M. E. Wrolstad, and H. Bouma (Eds.), Processing of visible language 2. New York: Plenum.

Wertheimer, M. (1958). Principles of perceptual organization. In D. Beardslee and M. Wertheimer, Eds., Readings in perception. Princeton: Van Nostrand.
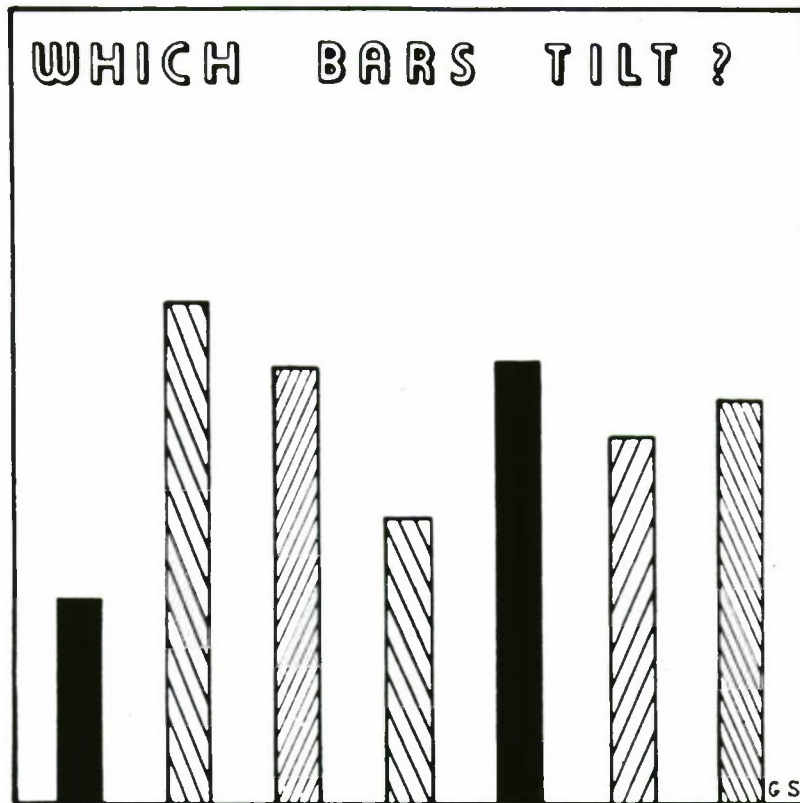
Figure 1.– Hypothetical graph taken from Schultz, G. M. (1961). Beware of diagonal lines in bar graphs. Prof. Geogr., 13, 28-29 (reprinted by Kruskal (1982)).
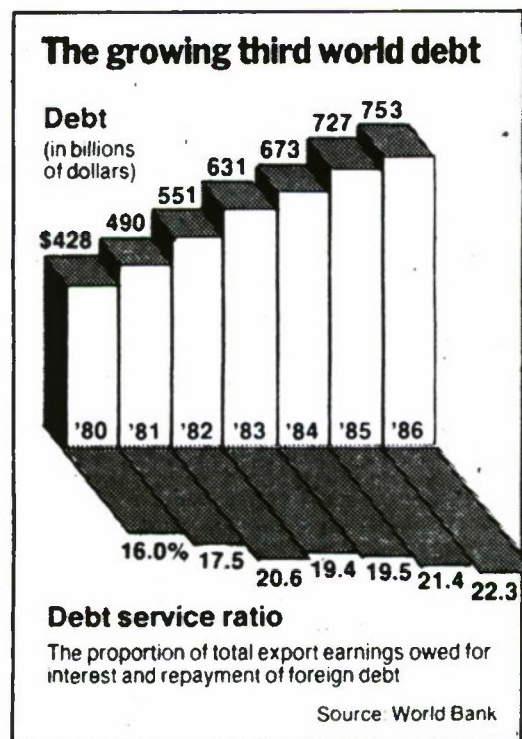


Figure 2.– Graph taken from The New York Times, August 2, 1987. (Copyright © 1987 by The New York Times Company. Reprinted by permission.)
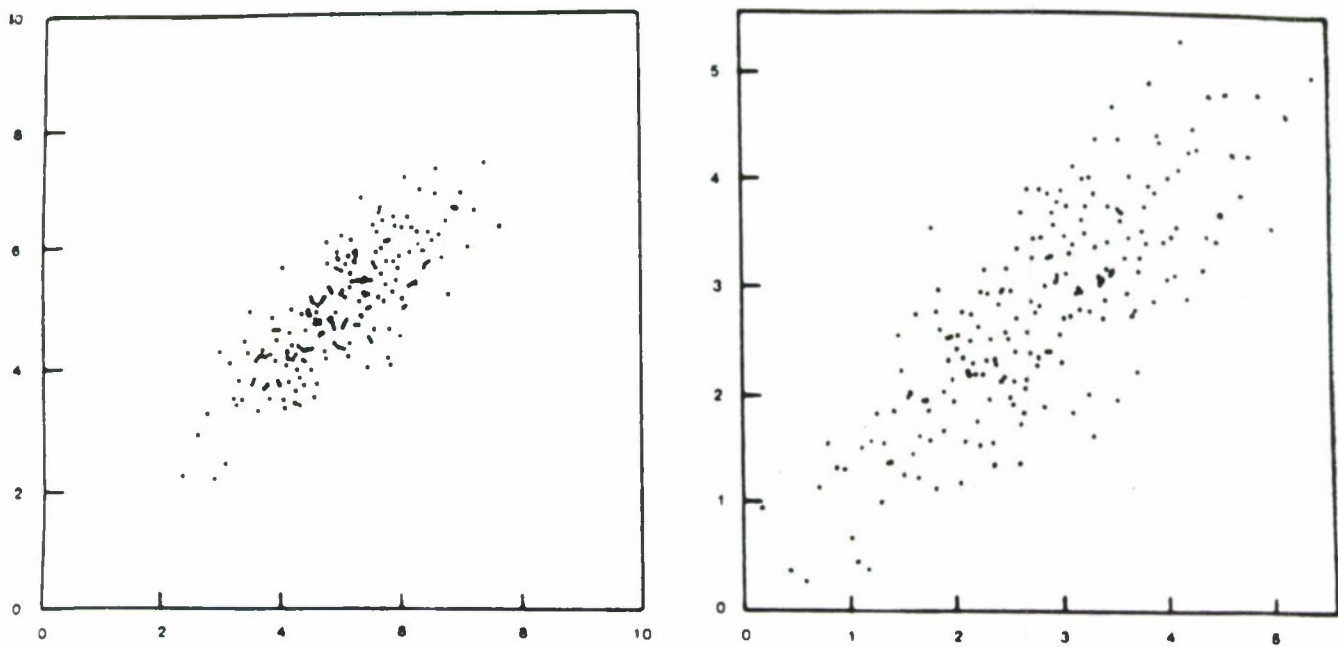
Figure 3.– Stimuli used by Cleveland, Diaconis, and McGill (1982). Although the correlations in the two scatterplots are the same, the right-hand one in the smaller frame is judged to be higher.
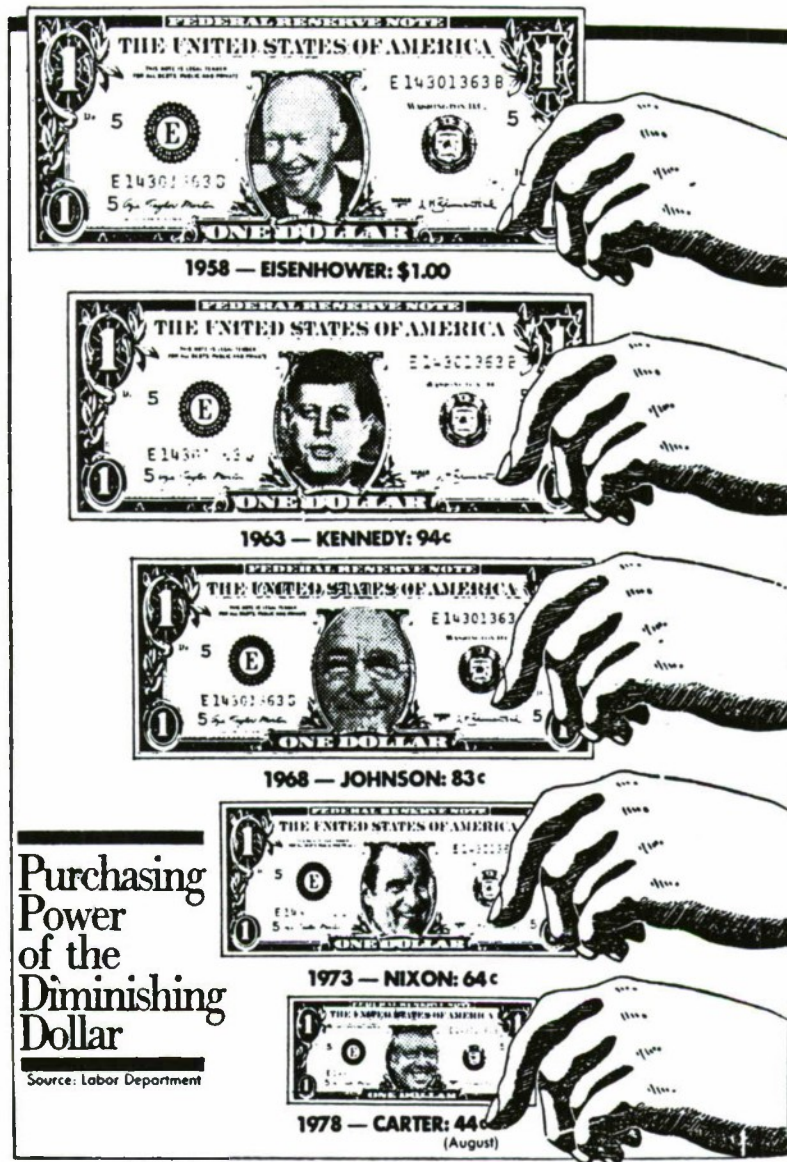
Figure 4.– Graph taken from The Washington Post, October 25, 1978 (reprinted by Tufte (1983) and Wainer (1980)).
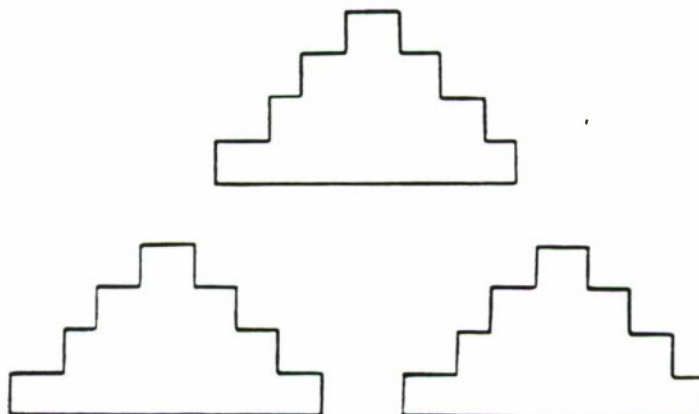


Figure 5.– Figures used by Freyd and Tversky (1984).

Figure 6.– Face taken from Neville (1977). The left and right faces were constructed by taking the left and right halves of the original photograph and reproducing them in mirror image, producing faces that are symmetric, unlike the original.
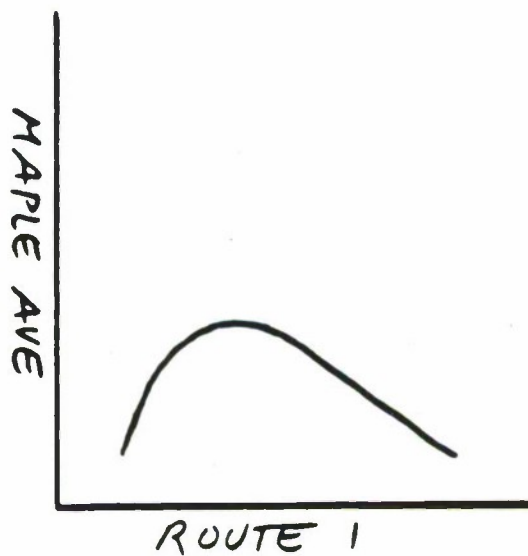


Figure 7.– Map curve used by Tversky and Schiano (1987 manuscript).
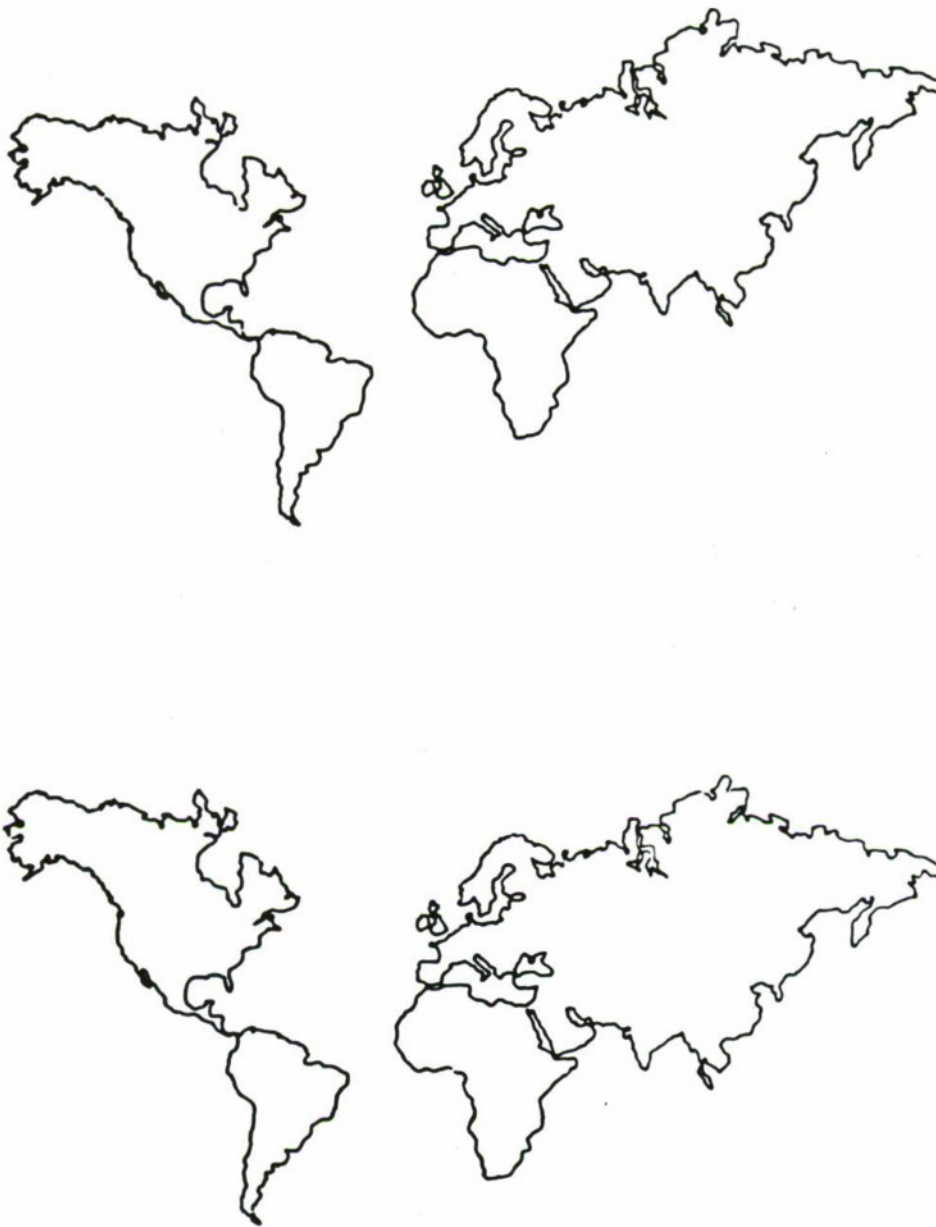
Figure 8.– World map stimuli used by Tversky (1981). Subjects incorrectly prefer the lower map, in which the U. S. and Europe, and South America and Africa are more aligned.
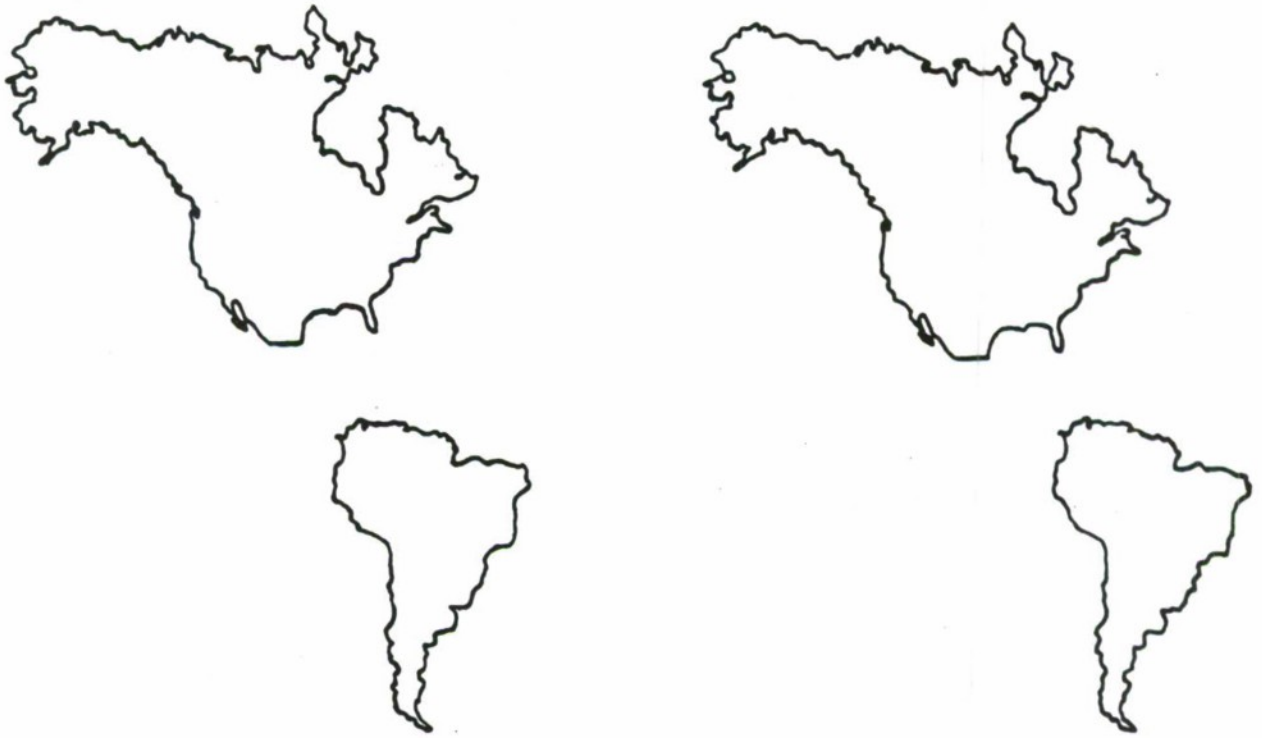
Figure 9.– Map of the Americas used by Tversky (1981). Subjects prefer the incorrect left one.
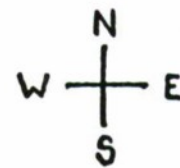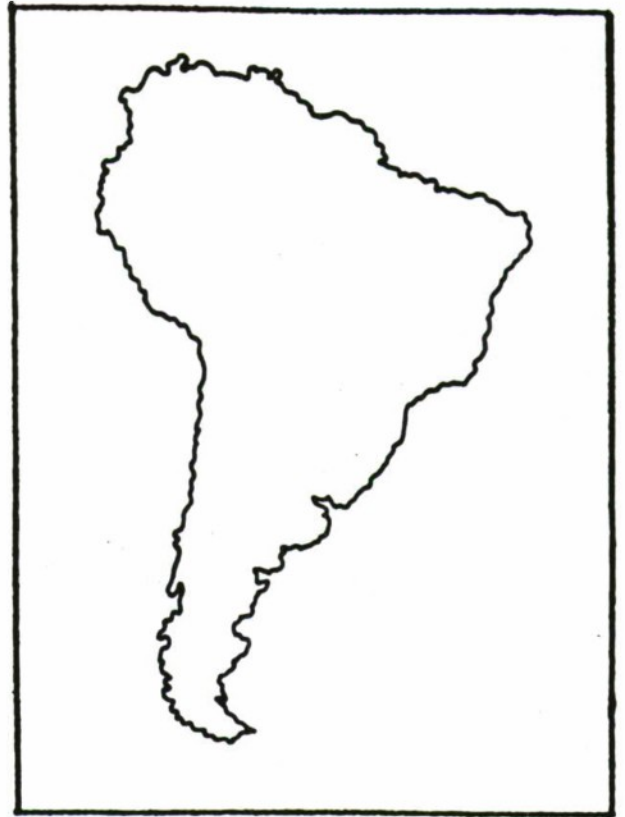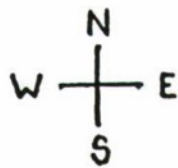
Figure 10.– The correct orientation of South America is on the right, but subjects typically upright it, as in the example on the left (from Tversky, 1981).
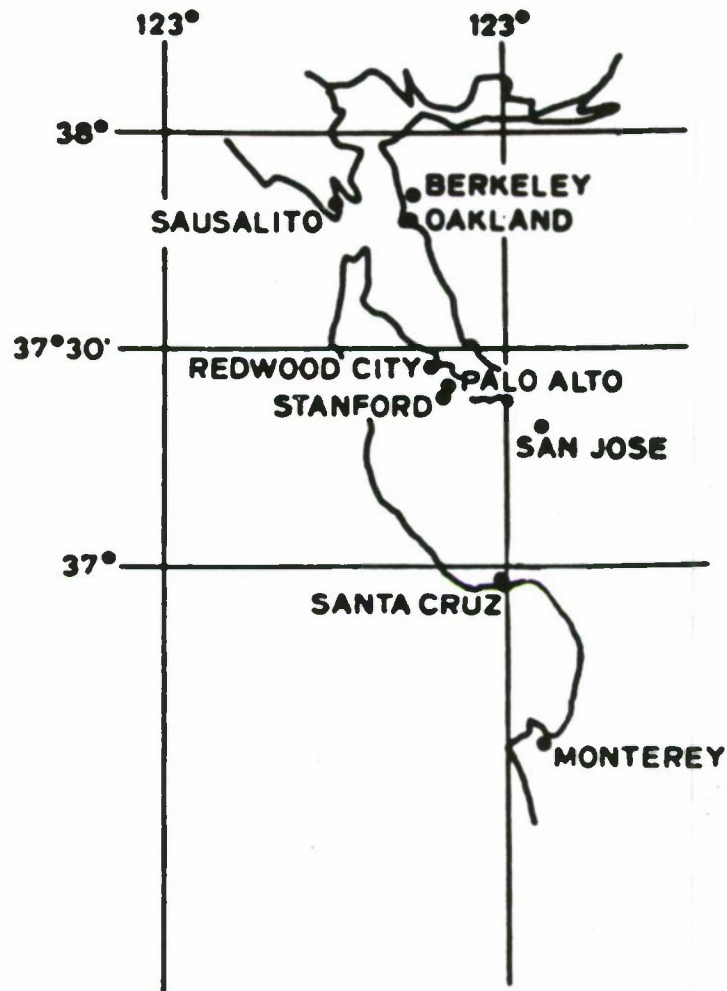
Figure 11.– The correct map of the San Francisco Bay area. Subjects erroneously report that Berkeley is east of Stanford and Palo Alto is east of Monterey (from Tversky, 1981).

| Process | Perceptual Cue | Perceptually-induced Error | Conceptually-induced Error |
|---|---|---|---|
| Figure segregation/ identification | Symmetry | toward symmetry | description enhances or reduces effect |
| Figure location/ orientation | Other figure | toward other figure | |
| | Frame of reference | toward frame of reference | graphic medium (e.g., maps vs. graphs) determines frame of reference (e.g., vertical/horizonal axes vs. 45° line) |

Figure 12.– Summary of perceptual analysis and resultant errors.
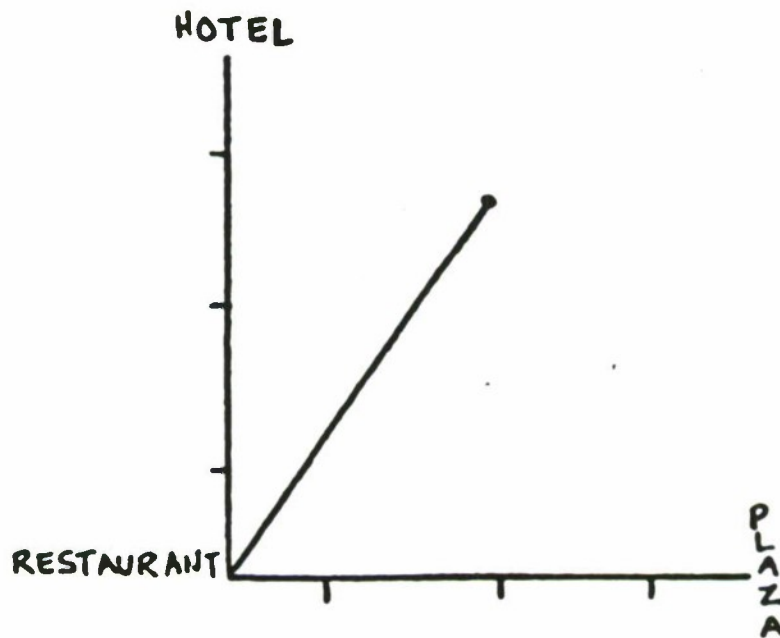
Figure 13.– Straight-line maps and graphs used by Tversky and Schiano (1987 manuscript).

# HELMET-MOUNTED PILOT NIGHT VISION SYSTEMS: HUMAN FACTORS ISSUES

Sandra G. Hart
Aerospace Human Factors Research Division

Michael S. Brickner
National Research Council, Research Associate
NASA Ames Research Center
Moffett Field, California

## ABSTRACT

Helmet-mounted displays of infrared imagery (forward-looking infrared (FLIR)) allow helicopter pilots to perform low-level missions at night and in low visibility. However, pilots experience high visual and cognitive workload during these missions, and their performance capabilities may be reduced. Human factors problems inherent in existing systems stem from three primary sources: (1) the nature of thermal imagery, (2) the characteristics of specific FLIR systems, and (3) the difficulty of using a FLIR system for flying and/or visually acquiring and tracking objects in the environment. The pilot night vision system (PNVS) in the Apache AH-64 provides a monochrome, 30° by 40° helmet-mounted display of infrared imagery. Thermal imagery is inferior to television imagery in both resolution and contrast ratio. Gray shades represent temperatures differences rather than brightness variability, and images undergo significant changes over time. The limited field of view, displacement of the sensor from the pilot's eye position, and monocular presentation of a bright FLIR image (while the other eye remains dark-adapted) are all potential sources of disorientation, limitations in depth and distance estimation, sensations of apparent motion, and difficulties in target and obstacle detection. Insufficient information about human perceptual and performance limitations restrains the ability of human factors specialists to provide significantly improved specifications, training programs, or alternative designs. Additional research is required to determine the most critical problem areas and to propose solutions that consider the human as well as the development of technology.

## INTRODUCTION

In most civil and military operations, helicopter pilots rely on visual cues to maintain situational awareness (e.g., estimate the orientation, altitude, speed, and direction of their vehicle; the location of hazards in the environment; and their geographical location). Maintaining visual contact with the environment is particularly important (and difficult) in nap-of-the-earth (NOE) flight, where pilots fly at altitudes between 10 and 30 ft, navigating in and among trees, hills, and buildings. During NOE flight, pilots must keep their eyes "out of the cockpit," rather than focused on displays within the cockpit. There is little margin for error. Existing electronic display systems do not provide adequately detailed information for visual flightpath control, and guidance algorithms do not yet exist for automatic NOE flight.

At night and in low visibility, the problem is more severe. Sufficient visual information about the environment is not available for pilots to navigate safely or identify relevant objects. For this reason, light-intensifying goggles and helmet-mounted displays of infrared imagery have been developed. This paper will focus on the unique visual environment created by the latter, as helmet-mounted displays of infrared imagery (alone or in combination with other sources of visual information) are integral to the design of many advanced helicopters.

Forward-looking infrared (FLIR) systems provide pilots with a monochromatic video image of the outside scene constructed from thermal differences among environmental features. Computer-generated flight symbology may be superimposed on the helmet-mounted display of FLIR imagery. Current FLIR pilot night vision systems (PNVS) can be used at night, in total darkness, or during the day, to allow pilots to "see" through blowing dust, smog, smoke, or concealing foliage.

The FLIR systems used in the Cobra AH-1S and the Apache AH-64 are turret-mounted on the nose of the helicopter. Their movement is slaved to the position of the pilot's helmet, allowing the pilot to move the 30° (vertical) by 40° (horizontal) instantaneous field of view (FOV) through a "field of regard" of ±90° in azimuth and 65° in elevation (from +20° to -45°) (fig. 1). The infrared sensor consists of an array of 180 detectors which provides 360 lines of resolution. This information is transformed into a 875-line video image which is displayed on a 1.92-cm combining lens (a monocle) mounted on the helmet immediately in front of the pilot's right eye. (fig. 2)

Given the integral role such systems are playing in advanced rotorcraft, it is surprising how little is known about human factors problems which are related to the use of these complex and highly demanding systems. The problems may be divided into three categories: (1) the unique nature of infrared images, (2) specific characteristics of the PNVS, and (3) problems related to the task of flying a helicopter at low altitudes in low-visibility conditions. This paper will focus on the most critical problem areas and evaluate their effects on pilot perception and performance.

## CHARACTERISTICS OF THERMAL IMAGES

Thermal images are a visible representation of radiation in the infrared band (8-14 μm in the PNVS). Thermal radiation is detected by an array of 180 detectors, in current-technology systems, which can create a visual display with approximately 360 lines of horizontal resolution. The output of each detector is preamplified, entered into a scan converter, transformed into a video image, and displayed on a combining lens mounted on the pilot's helmet.

The temperature of an object depends on the properties of its component materials and on its exposure to natural or artificial sources of heat. Its "thermal signature" depends primarily on its heat-emitting characteristics. The quality of a thermal image depends on the thermal signatures of terrain features and objects; the presence of thermal variability in the environment and atmospheric conditions (e.g., ambient temperatures, moisture, dust, and haze); and the sensitivity and size of the detectors. Current systems have a limited bandwidth which acts as a low-pass filter, effectively limiting the detail with which objects can be depicted.

Since FLIR images are transformed into video images and displayed on a cathode ray tube (CRT), they inherently suffer from all of the shortcomings of video imagery (e.g., limited

resolution, restricted contrast sensitivity, and dynamic brightness range). In addition, they are displayed monochromatically and provide a two-dimensional representation of the three-dimensional world. In comparison to video images, the display provided by the PNVS is also subject to the specific properties of FLIR technology and the unique characteristics of the thermal (as compared to the visual) properties of objects in the environment. Figure 3 depicts an example of a FLIR image with superimposed symbology.

The meaning of "bright" and "dark" in the thermal image is not necessarily equivalent to light and shade in the optical sense. An object may emit little heat because it is shaded, or for a variety of other reasons related to the nature of the material and its "thermal history" (Lloyd, 1975). Thus, in a given image, there may be "shades" which are partly equivalent to real optical shades, or there may be no shading whatsoever. The human eye has been trained to interpret dark spots as shaded areas. These are usually perceived as low spots or valleys in the terrain. Thus, pilots may try (inappropriately) to impose the same perceptual rules on thermal images. Furthermore, the brightness of a displayed object does not provide accurate range information because objects which emit high thermal energy may appear to be closer than they really are. Such misinterpretations of the terrain structure may have severe consequences for helicopter flight at very low altitudes.

The relative temperature of an object changes because of ambient temperature, internal heat production, and its heat-emitting characteristics. Thus, its infrared signature may change dynamically over time. Further, when the temperature of the "foreground" and "background" are near the same value (e.g., the "crossover" point) an object may disappear from the visual display. For example, a truck on a snow-covered field would be quite visible while its engine is running, but virtually invisible after sitting with its engine off for several hours. There are relatively predictable periods during each day when the temperatures of specific substances are very nearly equal. For example, water and vegetation may have two crossover points each day, under some conditions (fig. 4). When crossover occurs, the ability of a FLIR system to discriminate is severely degraded. The net result is very poor image quality (Berry et al., 1984).

During the day or soon after sunset, there may be high thermal contrasts, depending on the terrain and on atmospheric conditions. When this occurs, there are wide temperature gradients, which generate clear and highly detailed images. Later in the night, thermal contrasts gradually diminish and images become less detailed. In addition, the effect of solar thermal radiation on the temperatures of different substances varies and elements of terrain features may cool at different rates during the night. For example, leaves cool more rapidly than branches. Thus, late at night trees may look as if they have shed their leaves because their temperature approaches that of the ambient air temperature. It may be quite confusing for a pilot to pass a grove of fully-leaved trees on the way to a mission and a grove of apparently dormant trees on the way back.

On the other hand, because of the chemical processes, leaves may emit their own heat. Thus, when the polarity of the system is set so that dark shades represent cooler objects, leaves are very bright in contrast to their dark appearance in optical images. These "blonde" trees seem to merge into the background, making it difficult for pilots to spot them from a distance. Such dynamic changes require pilots to use complex rules of thumb to interpret visual images, yet accurate evaluations are critical for pilots flying below treetop level.

Urban areas generate and accumulate considerable heat during the day, but, as they cool during the night, temperatures tend to equalize. This can make it virtually impossible for a pilot to identify a specific object (such as a high building) which would stand out in an optical image.

Human-made sources of thermal radiation, such as engines, fires, and friction, provide small, but significant, sources of infrared radiation. An operating truck, for example, might have a hot spot near the location of the engine and another near the wheels. Thus, the thermal "signature" of the truck is quite different from its optical image. Furthermore, if the truck remains stationary, with its engine off, it may become difficult to discriminate from the surrounding terrain. The changing visual appearance of human-made objects presents a particularly critical problem for military pilots performing target identification and tracking.

Because infrared detectors are sensitive to relative rather than to absolute temperatures, and because most FLIR sensors scan horizontally (parallel to the horizon), the horizon may blend with the ground and sky (Bohm, 1985). The absence of a clear horizon line may have a detrimental effect on spatial orientation and altitude estimation.

## Display Polarity

Pilots may elect to assign either light or dark values to "hot" objects in the environment. Depending on the circumstances, they may alternate between the two polarities, selecting the one that provides the clearest image. Unlike the difficulties that people encounter in interpreting negatives of optical images, pilots can often improve their ability to recognize objects and interpret terrain features by switching the polarity of the FLIR display. For example, the sky is usually perceived as a bright area in an optical image, and it is always colder than the terrain. Thus, when the polarity is set to white-cold, the sky will appear to be bright. However, this will coincidentally result in some shaded areas also appearing as bright areas, in contrast to everyday experience. Thus, under a specific set of circumstances, one polarity might provide the most interpretable image for targeting or geographical orientation, while the other might be optimal for pilotage.

## Gain and Level

The visual display may, at any given moment, present only a sample of the dynamic temperature range. "Gain" and "level" controls allow the pilot to select the desired range of displayed temperatures. A specific combination of gain and level may or may not be optimal for a particular task. For example, if gain and level are set to be very sensitive to temperature variations within hot target areas, an insufficient number of gray shades might be available to provide a detailed image of the general scene. Some advanced systems offer automatic control over gain and/or level, to provide an optimal presentation of the average range of temperatures, without requiring the pilot to make control adjustments. This solution, while intended to reduce pilot workload, may be suboptimal for detecting a specific object in a given setting.

In summary, thermal images have some unique characteristics that result from the nature of infrared radiation. Human perceptual skills, which provide efficient tools for interpreting the "optical world," may be misleading when applied to thermal images. Research is necessary to (1) determine how the unique characteristics of infrared imagery interact with various aspects of human performance, (2) define the skills that are necessary to use FLIR displays of thermal images, and (3) establish how such skills should be acquired.

# SPECIFIC CHARACTERISTICS OF THE PNVS

In addition to the inherent characteristics of infrared imagery, many of the human factors problems identified in current systems are related to specific components and design limitations of the PNVS itself.

## Sensor Location

In the Apache, the FLIR sensor is mounted 3.5 m in front and 1.2 m below the pilot's eye position, creating a displaced eyepoint (fig. 5). Thus, objects within the field of regard of the sensor may be physically closer to the sensor than they are to the pilots' natural visual reference (his eyes) (Berry et al., 1984). During training, pilots must learn to adapt to a different visual reference point and adopt slightly different rules of thumb for estimating range and altitude using the PNVS display. In addition, objects abeam the sensor (which are no longer visible on the monocle) might not have passed the pilot's natural visual reference point, creating the possibility of confusion if the object is also visible to the pilots' unaided eye (fig. 6).

Since the sensor is located closer to the ground than are the pilots' eyes, available visual motion cues indicate slightly higher apparent velocities than pilots would estimate with direct vision. Again, during training, they must learn new rules of thumb to estimate their speed using the PNVS display. The displaced eyepoint creates motion parallax problems which are particularly severe when large viewing azimuths are encountered.

## Sensor Movement

In the Apache, the FLIR sensor responds to pilot head movements, moving at a rate of approximately 150°/sec. However, the slight delay between movement of the helmet and movement of the sensor can contribute to motion parallax problems. Although pilots learn to limit the frequency and velocity of their head movements to reduce such problems, certain tasks may require both rapid and frequent changes in the orientation of the sensor to a specific location or object within the FOV of the sensor.

## Helmet-Mounted Display Unit

In the Honeywell Integrated Helmet and Display Sighting System (IHADSS) used in the Apache and the Cobra "surrogate trainer" (where some pilots are familiarized with the system), infrared imagery is displayed as a rectangular area on a combiner lens incorporated into the helmet-mounted display unit (HDU). The lens is a semitransparent viewing screen that filters light in the red and blue range and reflects the composite video image presented in the green wavelength. The back of the lens is chemically coated to reduce glare, transmitting 50% of the light incident upon it. The lens reflects 80% of the green light rays that exit the HDU toward the pilot's eye. The end result of the filtering, magnifying, collimating, and reflecting processes is a two-dimensional, monochromatic, monocular display with a maximum of 125-150 ft-L of brightness (Berry et al., 1984).

## Field of View

The image presented to the pilot by the PNVS/IHADSS represents a rectangular FOV of 30° by 40°. The pilot views an image which is equivalent to a 7-ft television screen viewed from a distance of 10 ft (Berry et al., 1984). This relatively narrow FOV eliminates peripheral information that is critical for visual flightpath control. In visual flight, pilots depend on peripheral motion cues to estimate speed and orientation and to develop a sense of object's structure from visual motion cues. In addition, pilots must maintain their awareness of significant terrain features, the position and identity of stationary objects, and the projected course of moving vehicles that surround them for navigation, tactical decision-making, and obstacle avoidance. However, the field of regard of the sensor limits pilots' abilities to maintain visual contact with objects that are located beside or behind their vehicle.

Surprisingly, little empirical information is available about pilots' FOV requirements for pilotage, navigation, and target acquisition or their performance capabilities with different FOV. Furthermore, the FOV requirements for a helmet-mounted PNVS are even less well-known. A pilot may be faced with the requirement to fly the vehicle while visually tracking a target moving off-axis to the direction of flight using the same helmet-mounted display as the primary source of visual information for both tasks.

Considerable effort is being devoted to providing a wider FOV in more advanced systems (up to 60° or 90°) or providing different sensitivity for the foveal and peripheral elements of such a display. However, it is not clear whether the additional cost will be justified by an improvement in performance. Even a 90° FOV does not provide all of the peripheral cues available to the unaided eye in good visibility. Furthermore, if the FOV is increased without also improving the resolution of the display, the result may be a wide, but inadequately resolved, view of the terrain.

## Display Resolution

Pilots have identified display resolution as one of the most critical problems in existing systems (Bennett and Hart, 1987), although the IHADSS provides 875 lines of display resolution. To some extent, the appearance of inadequate display resolution could reflect the fact that the image is presented in close proximity to the pilot's eye. For example, the panel-mounted PNVS display has the same resolution as the helmet-mounted version, but it is viewed from a greater distance. This creates the impression of better resolution.

In fact, the apparent limitations in display resolution reflect the capabilities of the entire system, rather than the quality of the display alone. The effective resolution of the PNVS is less than 360 horizontal scan lines. Thus, in a 30° vertical FOV each scan line covers 5-6 min of visual angle, as compared to the resolving power of the human eye of about 1 min of arc. This is a substantial limitation in the level of detail that is available for presentation by the display system. For example, pilots report having great difficulty in detecting wires or other small targets, unless their thermal contrast with the surrounding environment is very high.

## Display Contrast

Advanced infrared detectors are capable of detecting temperature differences of approximately 0.3°C (Haidn, 1985). And a high-quality CRT can display at least 64 shades of gray. However, the PNVS provides only 10 shades of gray (ranging from bright to dark) to represent temperature differences in the environment (Tucker, 1984). This limitation severely restricts the level of detail that can be displayed at any one time and may interact with other limitations (e.g., limited resolution) to produce an unacceptable image quality.

Furthermore, specific gain and level selections, which are intended to enhance contrast in one region of the total range, might limit detail in another. For example, if the system is set to provide maximum contrast between the extremes, discriminations in the midrange will be limited. Conversely, when the display is optimized to provide fine discriminations in the midrange, extreme thermal signatures may not be discriminable. Because of the restricted number of gray shades provided to depict an image, the tolerance for inappropriate gain and level settings is very limited.

## Monocular Presentation

At night, the image presented by the PNVS/IHADSS effectively limits peripheral vision in the right eye, because the display is so bright in comparison to the environment. However, a full monocular FOV is still available to the unaided left eye (although visible cues may be limited on a dark night). Certain details and distance judgments may be obtained more accurately with the unaided (left) eye than with the aided (right) eye. Thus, pilots must rely on both sources of visual information. However, under most circumstances, the same object viewed by both eyes cannot be merged into a coherent binocular image, because of the differences in brightness, perceived size, and perceived location (resulting from the displaced eyepoint of the sensor.) To make matters worse, the right eye may be adapted to the bright image provided by the PNVS/IHADSS system, while the left eye might be dark-adapted to the environment. The problem of motion parallax created by the displaced eyepoint provided by the sensor location is particularly great in good visibility (where the unaided eye receives a clear image).

In practice, the use of available visual cues to augment information provided by the sensor may create more of a handicap than a help, because of competition between images presented to the two eyes (binocular rivalry). One consequence of binocular rivalry is that the information available in one eye, by competing for pilot's visual attention, may partially or completely suppress information available to the other eye. Furthermore, since pilots are trained to use both eyes when flying with a PNVS, they must learn how to process disparate visual cues, or shift their attention between their right eye (to use the PNVS) and left eye (to view the terrain or panel instruments.) To some extent, the focus of visual attention is under the pilot's conscious control. However, pilots report increasing difficulty in controlling the focus of visual attention as missions progress. After less than 1 hr of continuous use, some pilots report they must close one eye (to restore the visibility of information in the other eye) or exert significant attentional effort (Bennett and Hart, 1987).

Shifting visual attention from one eye to the other (without closing the unattended eye) is difficult to learn, mentally demanding, and visually fatiguing. Operational experience does not appear to minimize the problem; rather, pilots learn how to minimize its impact on their operational performance. It is not clear whether specific training programs, developed to aid pilots in

developing visual-attention-management skills, would be effective in improving pilot's performance and in reducing visual fatigue.

## Depth Perception

Because information is presented monocularly, all stereoscopic depth cues for objects in the immediate environment are lost. Additionally, the difference between the apparent size and location of objects viewed directly or through the sensor can provide conflicting information about the distance of objects in the environment (Roscoe, 1987). Although binocular systems have been proposed by government and industry researchers, the technical problems associated with fusing information from two sensors to provide a natural binocular image have not been solved adequately for operational use. Alternatively, the same image could be presented to both eyes–a biocular display. While this would eliminate the problem of binocular rivalry, it would limit pilots' abilities to gain peripheral cues outside the cockpit, see instruments inside the cockpit, or maintain at least one dark-adapted eye. And, it would still not provide stereoscopic information.

## Display Magnification

The displayed information is collimated to optical infinity and magnified to represent a 1:1 mapping with respect to the environment. However, the apparent magnification is not perceived as being 1:1. This creates a problem when precise distance judgments must be made, as during landing or formation flying. Pilots report that objects appear to be closer when viewed through a FLIR than they would with the unaided eye, particularly when the FLIR image is very bright (Bennett and Hart, 1987). Other distance misperceptions may also result from the difference in light and dark adaptation of the aided and unaided eyes (the Pulfrich effect, see Tyler, 1974) and from misaccommodation of the eyes (Roscoe, 1985). Pilots have reported that they minimize this problem by confirming range with their left eye. This forces them to shift their visual attention back and forth between the aided (light-adapted) and unaided (dark-adapted) eyes (Bennett and Hart, 1987).

## Summary

Current technology systems provide pilots with a wealth of information that would not otherwise be available at night or in low visibility. Without visual aiding, the range of environments in which low-level missions could be performed would be severely reduced. However, many properties of existing systems (e.g., low resolution; the restricted scale of gray shades; and a limited, monocular field of view) contribute to the creation of images which contain only a small part of the information that is available through direct vision in good visibility. Thus, pilots are deprived of essential information about small obstacles or targets and the detail required to identify larger objects. The adverse effects of degraded image quality may impose significant workload and visual fatigue. However, the effects of these factors seem to be relatively unequivocal and predictable, in comparison with the effects of sensor location, binocular rivalry, and depth perception. These phenomena may appear in different forms during different flight maneuvers and for different pilots. Some individuals may even experience exactly the opposite phenomena than others experience. For example, some pilots tend to overestimate, while others underestimate, size

and distances. Thus, considerable skill and experience is required for NOE flight with the PNVS, and even highly trained pilots consider it to be a highly demanding task.

## ISSUES RELATED TO HELICOPTER CONTROL

In addition to all of the human factors problems related to the nature of the thermal image and to the design of the PNVS/IHADSS, one has to bear in mind that the system is installed on a moving, six-degree-of-freedom platform which is designed to perform a variety of demanding operational tasks. Some of the most difficult tasks involve NOE flight, off-axis tracking, and hovering.

To perform each of these tasks well, pilots must learn to distinguish the effects of control inputs (e.g., changes in the direction, speed, or orientation of the helicopter itself) from the effects on the visual display of changes in sensor orientation induced by the pilot's head movement. Disorientation can result from a conflict between vestibular cues (based on vehicle motion) and visual cues (obtained through the sensor). Pilots learn to limit their head movements (to reduce vertigo) and to time them to achieve a stable direction of gaze before changing their direction of flight (to reduce spatial disorientation.) They must balance this requirement for limited head movement against their need to scan the environment (to obtain an acceptable field of regard or to track moving targets) to compensate for the sensor's narrow FOV.

### NOE Flight

In NOE flight, pilots must fly at very low altitudes among natural and human-made terrain features. Even in good visibility, this presents a challenging task for which there is a very low tolerance for error. In reduced visibility, the requirement to perform the same mission using visual aids (such as the FLIR/PNVS) is even more difficult. In NOE flight, problems associated with the quality of the visual display, the absence of stereoscopic depth cues, display magnification, and the offset location of the sensor are particularly pronounced and combine to make rapid and accurate range estimates, required to avoid natural and human-made obstacles, very difficult. In addition, it is difficult for pilots to maintain a sense of their general geographical orientation because of the narrow FOV of the sensor and limitations in its range; their view of the world through which they are flying is effectively limited to nearby terrain features. Also, the degraded and dynamically changing quality of the visual representation of objects in the environment make it difficult for pilots to detect and recognize otherwise familiar objects and terrain features. Finally, the narrow FOV of the sensor and limitations in the display of surface texture inhibit pilots' abilities to maintain visual control of speed, heading, and altitude.

These limitations combine to create a flight environment where pilots must fly slower and higher to maintain acceptable margins for safety. Further, performing this task imposes high visual and cognitive demands on pilots and is very fatiguing, thereby limiting the duration of missions and flight hours.

## Hovering

In an inherently unstable vehicle, or without stability and control augmentation systems, hovering is extremely difficult and performance is worse when visual information is obtained through a helmet-mounted display (Landis & Aiken, 1982). Even in a relatively stable vehicle, such as the Apache, visual reference points vary whenever the pilot moves his or her head and depth cues are difficult to obtain from the monocular display. Because display resolution is limited, subtle relative motion cues may be difficult to detect. In addition, peripheral visual cues that provide an important source of motion information with direct vision are limited on the PNVS/ IHADSS. Thus, pilots supplement the sensor imagery with information available to the unaided eye (to provide the necessary peripheral motion cues) and with information provided by superimposed symbology.

## Off-axis Tracking

Since the sensor is attached to the helicopter, its orientation and position with respect to the environment reflect the forward, lateral, and vertical translation and pitch, roll, and yaw of the vehicle. However, within the boundaries of its range of movement, the azimuth and elevation of the FLIR sensor is independent of the orientation of the helicopter. Spatial disorientation and reduced flightpath control performance may occur when pilots look in a different direction than the vehicle is moving ("off-axis" tracking). Visual motion cues relevant for flightpath control are more difficult to interpret when they are obtained through a sensor that is oriented off-axis to the direction of flight (see fig. 6). Peripheral cues (which could integrate the conflicting sources of information) are limited by the narrow FOV, thereby intensifying the problem.

Pilots appear to trade off flight-control performance for visual tracking performance; visual tracking performance is degraded when it is coupled with the requirement to control the vehicle. In addition, visual tracking of curved vehicle trajectories is degraded (in comparison to straight trajectories) and tracking error is increased as the apparent rate of movement of a target across the pilot's visual field is increased (by changes in the distance of a target, the rate of movement of the target, and/or the velocity of the pilot's vehicle) (Bennett et al., this volume).

Pilots report (Bennett and Hart, 1987) that they are able to perform off-axis tracking for only short periods of time (no more than a few seconds, depending on the flight mode) before they must return the orientation of the sensor to correspond to the direction of flight. Thus, pilots come to a hover (when they must visually track a moving target) or they hand a target off to the copilot. Research is under way at NASA Ames Research Center (Bennett et al., this volume) and elsewhere, to quantify the range of human performance limitations in performing off-axis tracking and to develop display augmentations to improve pilots' performance capabilities.

## Superimposed Symbology

Several sources of information are often combined on helmet-mounted displays. In the Apache AH-64 and the Cobra, computer-generated symbology depicting flight-control information is superimposed on the sensor imagery and presented on the HDU. This composite display reduces the need for pilots to look at cockpit instruments during low-level flight.

**Flight-control Symbology**- In the Apache, computer-generated graphic and symbolic information about the vehicle's flight and performance status is provided to improve pilots' abilities to perform flightpath control. The computer-generated display is visible on the monocle no matter where the pilot's head points. However, since the symbology is always oriented in the direction of flight, as it would be in a head-up or panel-mounted display, it may not present the flight-control symbology in an orientation that is compatible with the direction the pilot is looking (fig. 7).

Up to 14 flight parameters may be displayed to ensure vertical and horizontal orientation. Different subsets of information are presented for different mission segments (e.g., hover, transition to hover) (fig. 8). The sensitivity of some elements of the display changes for different tasks (e.g., sensitivity is increased during hover and for given altitudes). Although such increased sensitivity is essential to allow pilots' to maintain a stable hover, learning how to interpret variations in the movement of symbolic display indicators is difficult during initial training (Bennett and Hart, 1987).

HDU displays of flight symbology are extremely useful, particularly in NOE flight when pilots are too busy to look at cockpit instruments. However, perceptual problems may be created by the interference between the computer-generated symbology (which is always oriented in the direction the vehicle is moving) and the video display upon which it is superimposed (which is oriented in the direction the pilot is looking) (see fig. 8 ). Furthermore, movement of the HDU symbology may induce a perception of apparent motion in the video display.

Pilots learn to ignore the superimposed indicators (when they do not need the information) to resolve the problem of display clutter. This is analogous to ignoring the dividers between panes of glass in a multipane window when looking outside—one only "sees" the outside scene. However, for windows, there is a difference in accommodation between the two sources of information, facilitating a difference in attentional focus. For the PNVS/IHADSS, on the other hand, the optical distance of both visual display elements is the same, increasing the difficulty that pilots have in focusing on one source of visual information or the other. Pilots report that they tend to look through the symbology at the outside scene (at the expense of viewing critical flight data) or vice versa (Bennett and Hart, 1987). When they feel that they do need the information, however, they include it in their scan. One symbol that remains essential is the diamond that represents the "nose" of the helicopter. It was added at the request of the first pilots to fly the PNVS to orient them to their direction of flight regardless of where the sensor was pointing.

**Targeting Information**- Weapons selection, aiming, and other targeting information can be superimposed on a display, as well. The Target Acquisition/Designation System (TADS) in the AH-64 provides FLIR, direct-vision optics, and daylight television display options boresighted to a common line of sight. The TADS has narrow and wide FOV alternatives and an electronic "zoom" capability. In the current configuration, the TADS is used by the copilot/gunner. However, in the environment envisioned for more advanced helicopters, such as the LHX, a single pilot might be required to use a helmet-mounted PNVS for both primary vehicle control and for weapons delivery. The visual display might be provided by one sensor or a fused combination of different sensors. In this situation, it is possible that a pilot might need to look in one direction to maintain vehicle control and in another to track, acquire, and fire at enemy targets. Command information might be displayed to tell pilots where to look if an automatic target recognition system

identified a target in a different direction than they were looking. This could result in a visual display of superimposed visual information from three different spatial orientations: (1) computer-generated symbology oriented in the direction of flight, (2) the display of FLIR information oriented in the direction of the pilot's head, and (3) targeting information.

## Effects of Vibration

Normally, the human eye is stabilized so as to maintain visual fixation in moving environments. The vestibular-ocular reflex induces eye movements that oppose those of the head to maintain a stationary point of regard during voluntary head movements. In vibrating environments, however, the eye may not be capable of compensating for the high-frequency components. The detrimental effects of vibration on visual acuity have been well documented (e.g., Griffin, 1977), particularly for panel-mounted displays, where some of the effects of vibration on instrument reading can be compensated for by presenting sufficiently large characters and symbols.

The effects of vibration can be even more severe with helmet-mounted displays, although the range of vibrations in advanced-technology helicopters has been reduced considerably. The sensor, which is slaved to the pilot's head movement, cannot discriminate involuntary, vibration-induced helmet movements from those initiated by the pilot. Relative motion is created between the image on the head-coupled display and the eye, resulting in retinal blurring, increased errors, and longer responses. It has been suggested that such "involuntary" head movements might be sensed by an onboard computer and that this information could be used to provide a stabilized display for the pilot (Velger, Grunwald & Merhav, 1986). Based on a computer simulation of the vibration frequencies of helicopters, an adaptive noise-canceling technique has been developed that minimizes the relative motion between viewed images and the eye by shifting displayed images in the same direction and magnitude as the induced reflexive eye movements. The filter stabilizes the images in space while still allowing low-frequency, voluntary head motions required for aiming accuracy.

## The Helmet

The IHADSS apparatus is relatively heavy (4 lb), producing discomfort and fatigue. And most of the weight is in front; counter-balancing weights do not completely eliminate the muscle fatigue induced by maintaining heads-up attention to the visual scene. In addition, to reduce the problems associated with involuntary head motion within the helmet, a snug fit is essential, which may produce "hot spots," further increasing discomfort. However, the pilots' helmets rarely fit perfectly with the consequence that the position of the monocle, which is attached to the helmet, may shift in flight. Furthermore, pilots' head movements within an imperfectly fit helmet may not be directly translated into helmet movements (which actually control the orientation of the sensor), although this does not present a major problem with current systems.

## Crew Size

All contemporary military helicopters have a flight crew of at least two. In attack helicopters one crew member is primarily responsible for flying the vehicle, while the other is responsible for navigation, target selection, and weapon control. Recently, the U.S. Army considered the

possibility of fielding a single-pilot helicopter. If a single pilot was required to perform a typical Apache mission, he or she would have to simultaneously control the helicopter during demanding flight maneuvers (e.g., NOE, hover) while detecting, acquiring, and destroying targets. It is well established in the motor-control literature that the concurrent performance of any two nonsynchronized motor tasks is extremely demanding and very difficult (e.g., Keele, 1986). Thus, effective off-axis target tracking seems to be feasible only if manual flightpath control demands are low (as in high-altitude, straight-and-level flight) or if at least one of the tasks can be automated. Since the high-threat battlefield environment requires NOE flight, automated flight and hovering systems may be required to effectively release a single pilot from the control of the platform (to enable the pilot to accomplish the weapons delivery task), or effective automated target recognition/ acquisition systems will be required to provide the pilot with reserve capacity to perform manual flightpath control. The successful design of a single-pilot, multipurpose helicopter will rely on the accumulation of a considerable body of human factors data in the areas of human information processing, workload, motor control, perception, and skill acquisition.

## Summary

Helmet-mounted pilot night-vision systems do what they are intended to do. They allow pilots to perform NOE missions at night and under low-visibility conditions. They do so at a considerable cost to the pilots, however, and adequate training can provide only a partial solution.

Current PNVS/IHADSS systems provide pilots with a monocular display of monochrome video images with limited resolution. The detector is not sensitive to natural variations in shading in the terrain and provides a narrow FOV from a displaced visual eyepoint. The appearance of thermal images may deviate substantially from optical images, and it changes with environmental conditions. The quality of the displayed image is further affected by (1) the existence of thermal contrasts in the environment; (2) the number of gray shades with which the sensor represents temperatures differences; (3) atmospheric conditions; (4) the selected polarity, gain, and level; and (5) vibration. Finally, there are additional limitations created by the display system itself (e.g., the resolution of the CRT and its monocular format).

These and other characteristics of current technology systems combine to provide pilots with limited visual cues under many circumstances. This, in turn, inhibits their ability to fly as low or as quickly as they might with optimal visual information. Some of the specific perceptual and cognitive problems that might contribute to such limitations in performance are (1) binocular rivalry (due to the monocular mode of presentation); (2) inaccurate range estimation (due to the offset sensor location); (3) loss of peripheral motion cues (due to the narrow FOV); (4) loss of directional orientation during off-axis tracking; (5) difficulty in identifying objects (due to limited display resolution and contrast and the unique properties of thermal images); and (6) loss of geographical orientation (due to the narrow FOV and limitations in the line of sight created by terrain features that obscure forward vision during NOE flight). Fatigue, especially visual fatigue, presents a particularly severe problem. And all of the issues discussed above may limit pilots' confidence in their ability to control their aircraft at low altitudes where misinterpretation of the structure of the terrain may have severe consequences. Finally, in addition to the operational limitations reported by experienced pilots, significant problems have been reported during training.

Although alternative designs have been suggested, there is insufficient information about human perceptual and performance limitations (and their interactions) to provide significantly

improved specifications, training programs, or alternative designs. Additional research is required to determine the most critical problem areas and to propose solutions that consider the human as well as the development of technology. Even though critical human factors problems with night-vision systems have already been identified, relatively little research is currently being conducted.

# REFERENCES

Bennett & Hart (1987) PNVS related visual problems: Pilots' reflections on visually coupled systems, 1976-1987. Working paper. Moffett Field, CA: NASA Ames Research Center.

Bennett, C. T. Johnson, W. W., Perrone, J., & Phatak, A. (This volume) Synthetic perspective optical flow: Influence on pilot control tasks. In Proceedings of the Spatial Displays and Spatial Instruments Conference and Workshop. Washington, D. C.: National Aeronautics and Space Administration.

Berry, J., Dyer, R., Park, R., Sellers, A., & Talton, M. (1984) PNVS Handbook Fort Rucker, AL: U.S. Army Training and Doctrine Command.

Bohm, H. O. V. (1985) FLIR, NVG, and HMS/D systems for helicopter operation - Review. In Visual Protection and Enhancement (AGARD CP-379) Neuilly-sur-Seinne, FR: Advisory Group for Aerospace Research and Development, 2-1 - 2-27.

Griffith (1977) The evaluation of human exposure to helicopter vibration. Aeronaut. J., March, 110-123.

Haidn, H. (1985) Operational experiences with night vision goggles in helicopter low level flight at night. In Visual Protection and Enhancement (AGARD CP-379) Neuilly-sur-Seinne, FR: Advisory Group for Aerospace Research and Development, 3-1 - 3-8.

Keele, S. W. (1986) Motor control. In K. Boff, L. Kaufman, and J. Thomas (Eds) Handbook of Perception and Human Performance, Vol II. New York: John Wiley and Sons. 30-1 - 30-60.

Landis, K. H. & Aiken, E. (1982) An assessment of various side-stick controller/stability and control augmentation systems for night nap-of-the-earth flight using pilot simulation. Paper presented to the American Helicopter Society, Palo Alto, CA.

Lloyd, J. M. (1975) Thermal Imaging Systems. New York: Plenum Press.

Roscoe, S. N. (1985) Bigness is in the eye of the beholder. Human Factors, 27(6), 615-636.

Roscoe, S. N. (1987) The trouble with HUDs and HMDs. Human Factors Bull., July, 1-3.

Tucker, G. E. (1984) Some thoughts on the implementation of pilot night vision devices for helicopters. Proceedings of the Technical Workshop: Advanced Helicopter Design Concepts. NASA CP-2351.

Tyler, C. (1974) Stereopsis in dynamic visual noise. Nature, 250, 781-782.

Velger, M., Grunwald, A. & Merhav, S. (1986) Adaptive filtering of biodynamic stick feedthrough in manipulation tasks onboard moving platforms. Technical Paper, AIAA Conference.
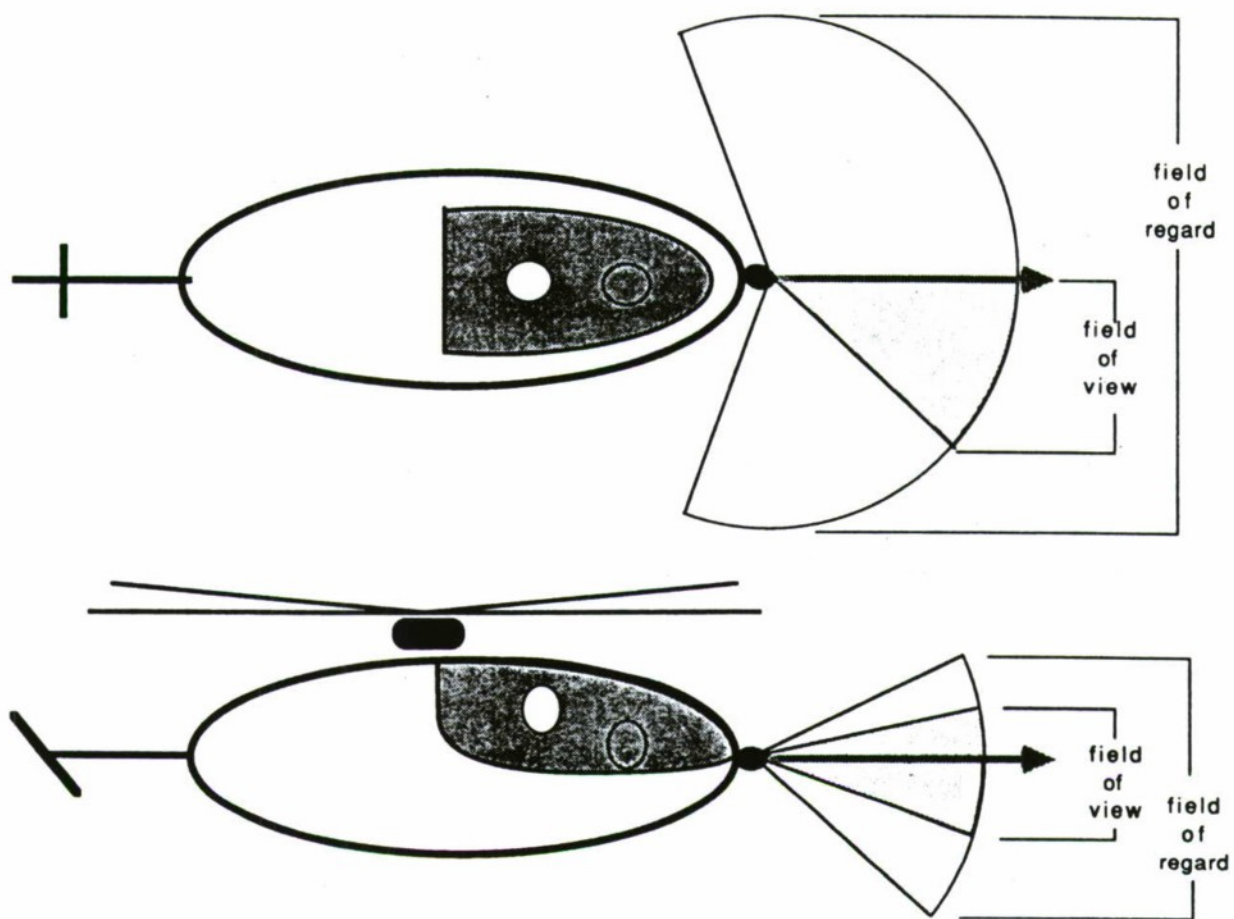
Figure 1.- Diagram of the vertical and horizontal FOV and fields of regard of the FLIR sensor.

Figure 2.- PNVS helmet-mounted display unit .
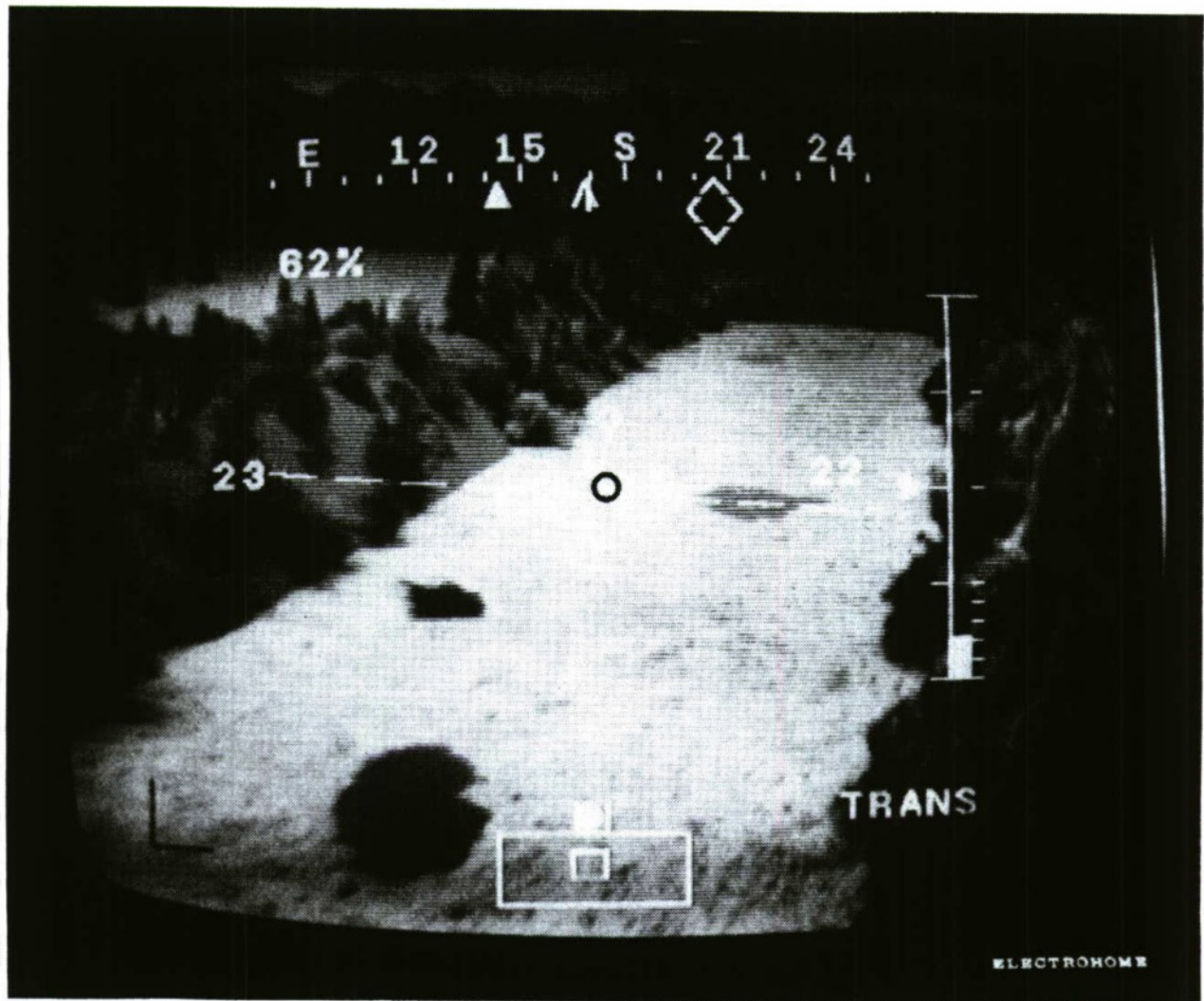
Figure 3.- Thermal display with superimposed flight-control symbology.
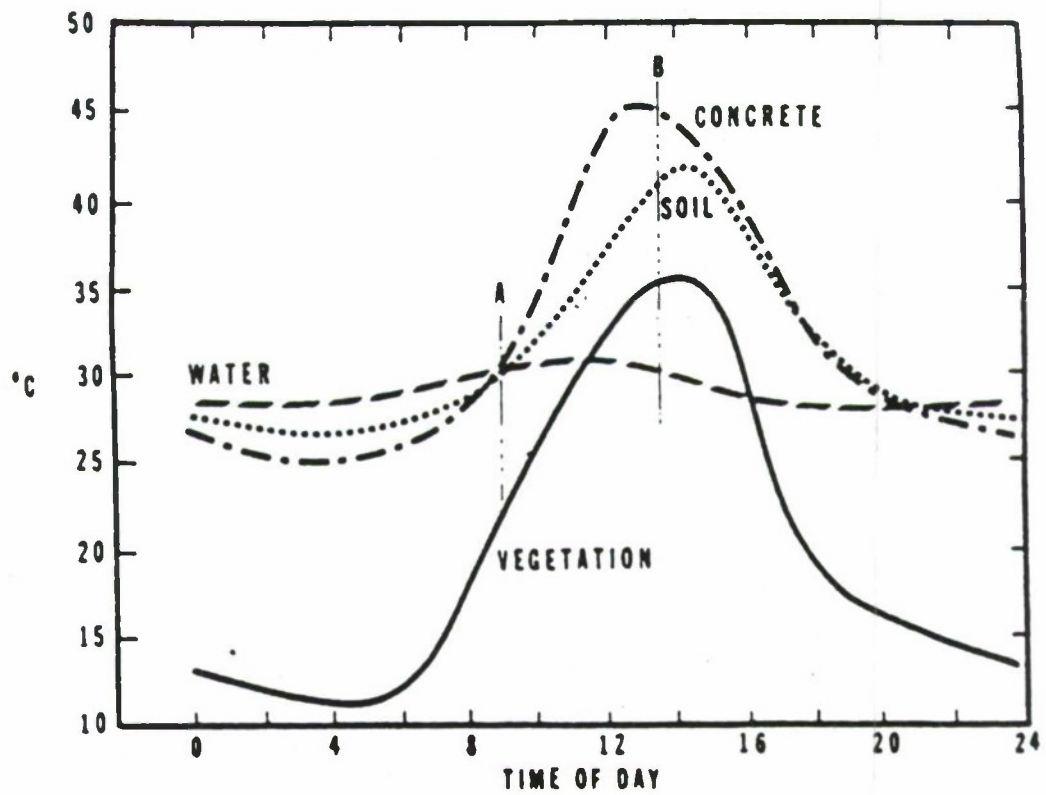
Figure 4.- Temperature distributions of different materials during a 24-hr period ("a" indicates the occurrence of crossover; "b" the time of day when the temperature differences are greatest) (Berry et al., 1984).
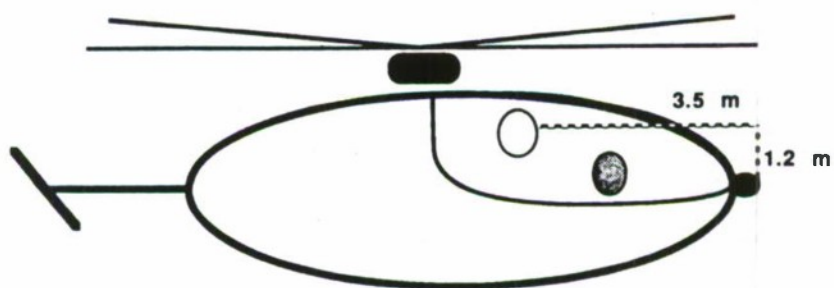


Figure 5.- Sensor offset.
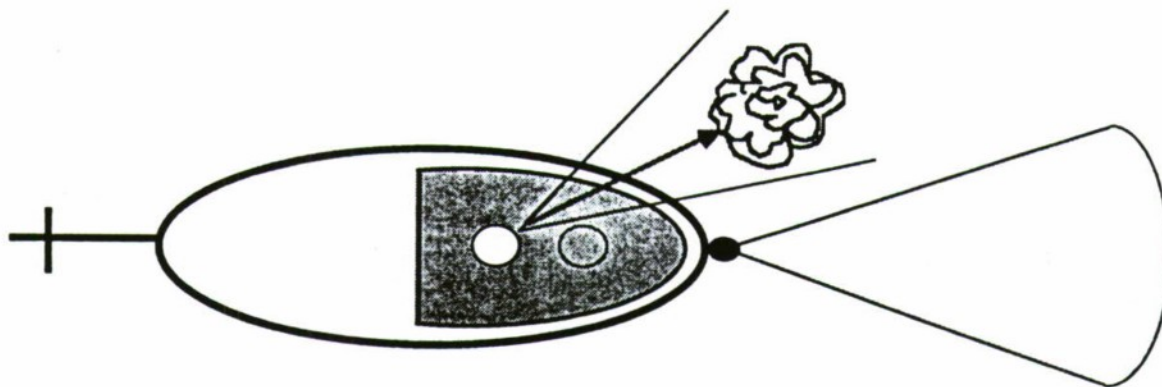
Figure 6.- Example of a situation where an object (a tree) seen by the pilot's unaided eye has passed behind the FOV of the sensor.
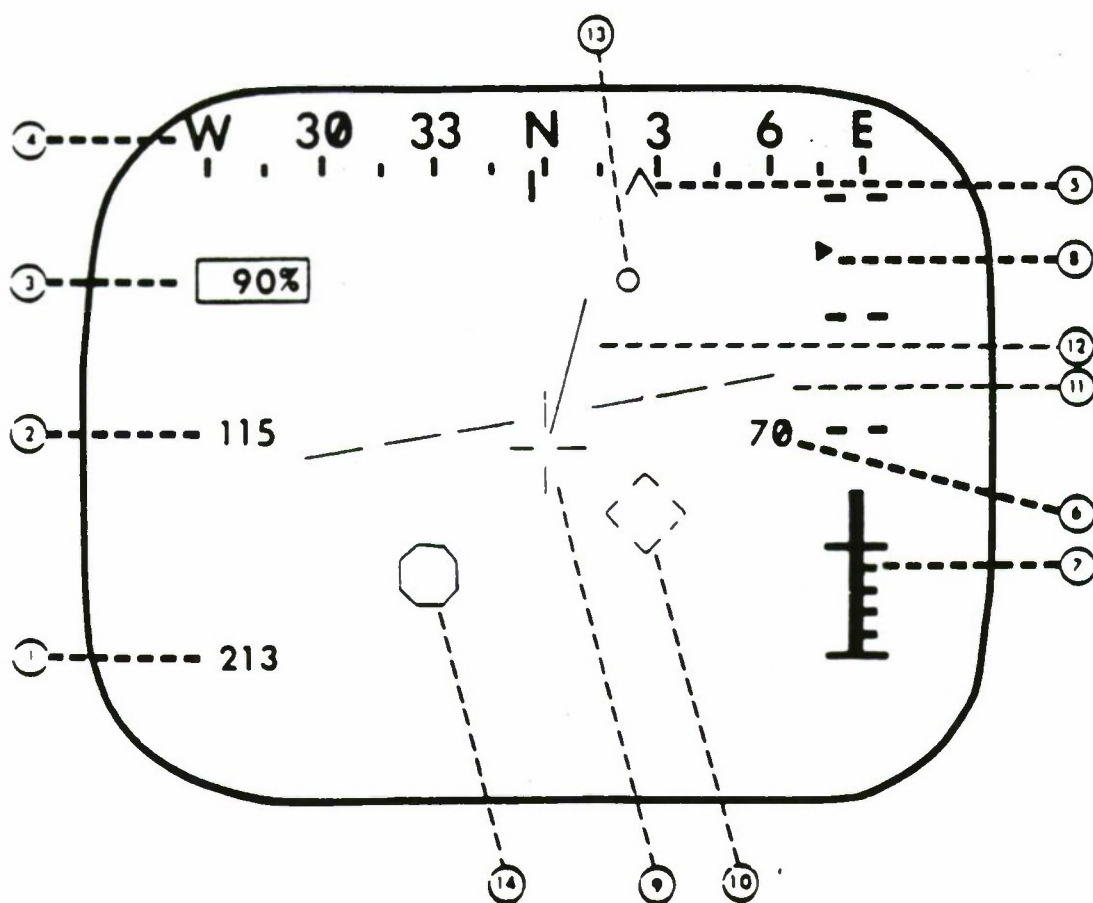


1--Digital ground speed
2--Digital indicated airspeed
3--Digital torque
4--Magnetic heading
5--Doppler steer indicator

6--Digital altitude
7--Analog altitude
8--Vertical speed indicator (VSI)
9-14--Central symbology

Figure 7.- Stylized example of different spatial orientations for FLIR imagery and superimposed computer-generated symbology.
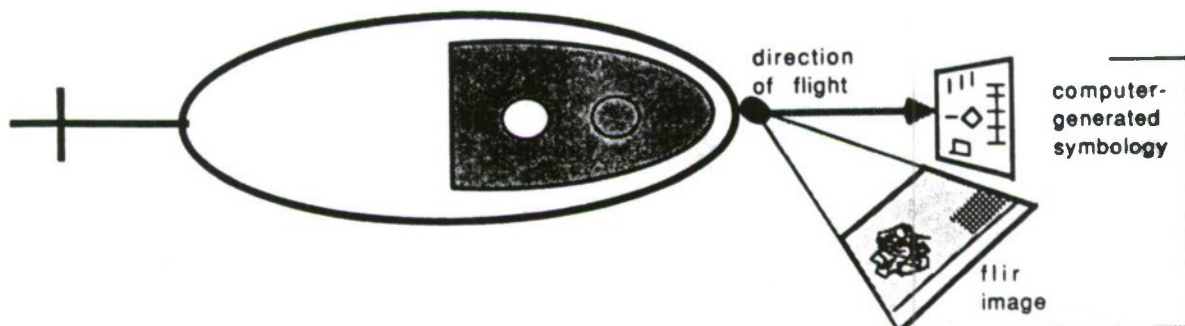
Figure 8.- The complete PNVS symbology set.

# SEPARATE VISUAL REPRESENTATIONS FOR PERCEPTION AND FOR VISUALLY GUIDED BEHAVIOR

Bruce Bridgeman
Department of Psychology
University of California
Santa Cruz, California

## SUMMARY

Converging evidence from several sources indicates that two distinct representations of visual space mediate perception and visually guided behavior, respectively. The two maps of visual space follow different rules; spatial values in either one can be biased without affecting the other. Ordinarily the two maps give equivalent responses because both are veridically in register with the world; special techniques are required to pull them apart. One such technique is saccadic suppression: small target displacements during saccadic eye movements are not perceived, though the displacements can change eye movements or pointing to the target.

A second way to separate cognitive and motor-oriented maps is with induced motion: a slowly moving frame will make a fixed target appear to drift in the opposite direction, while motor behavior toward the target is unchanged. The same result occurs with stroboscopic induced motion, where the frame jumps abruptly and the target seems to jump in the opposite direction.

A third method of separating cognitive and motor maps, requiring no motion of target, background or eye, is the "Roelofs effect": a target surrounded by an off-center rectangular frame will appear to be off-center in the direction opposite the frame. Again the effect influences perception, but in half of our subjects it does not influence pointing to the target. This experience also reveals more characteristics of the maps and their interactions with one another—the motor map apparently has little or no memory, and must be fed from the biased cognitive map if an enforced delay occurs between stimulus presentation and motor response.

In designing spatial displays, the results mean that "what you see isn't necessarily what you get." Displays must be designed with either perception or visually guided behavior in mind.

The visual world is represented by several topographic maps in the cortex (Van Essen, Newsome, and Bixby, 1982). This characteristic of the visual system raises a fundamental question for visual physiology: do all of these maps work together in a single visual representation, or are they functionally distinct? And if they are distinct, how many functional maps are there and how do they communicate with one another? Because these questions concern visual function in intact organisms, they can be answered only with psychophysical techniques. This paper presents evidence that there are at least two functionally distinct representations of the visual world in normal humans; under some conditions, the two representations can simultaneously hold different spatial values. Further, we are beginning to understand some of the ways in which the representations communicate with one another.

Experiments in several laboratories have revealed that subjects are unaware of sizeable displacements of the visual world if they occur during saccadic eye movements, implying that information about spatial location is degraded during saccades (Ditchburn, 1955; Wallach and Lewis, 1965; Brune and Lücking, 1969; Mack, 1970; Bridgeman, Hendry, and Stark, 1975). Yet people do not become disoriented after saccades, implying that spatial information is maintained. Experimental evidence supports this conclusion. For instance, the eyes can saccade accurately to a target that is flashed (and mislocalized) during an earlier saccade (Hallett and Lightstone, 1976), and hand-eye coordination remains fairly accurate following saccades (Festinger and Cannon, 1965). How can the loss of perceptual information and the maintenance of visually guided behavior exist side by side?

To begin a resolution of this paradox, we noted that the two kinds of conflicting observations use different response measures. The saccadic suppression of displacement experiments require a nonspatial verbal report or button press, both symbolic responses. Successful orienting of the eye or hand, in contrast, requires quantitative spatial information. The conflict might be resolved if the two types of report, which can be labeled as cognitive and motor, could be combined in a single experiment. If two pathways in the visual system process different kinds of information, spatially oriented motor activities might have access to accurate position information even when that information is unavailable at a cognitive level that mediates symbolic decisions such as button pressing or verbal response. The saccadic suppression of displacement experiments cited above address only the cognitive system.

In our first experiment on this problem (Bridgeman et al., 1979), the two conflicting observations (saccadic suppression on one hand and accurate motor behavior on the other) were combined by asking subjects to point to the position of a target that had been displaced and then extinguished. Subjects were also asked whether the target had been displaced or not. Extinguishing the target, and preventing the subjects from viewing their hands (open-loop pointing), guaranteed that only internally stored spatial information could be used for pointing. On some trials, the displacement was detected, while on others it went undetected, but pointing accuracy was similar whether the displacement was detected or not.

This result implied that quantitative control of motor activity was unaffected by the perceptual detectability of target position. But it is also possible (if a bit strained) to interpret the result in terms of signal detection theory as a high response criterion for the report of displacement. The first control for this possibility was a two-alternative, forced-choice measure of saccadic suppression of displacement, with the result that even this criterion-free measure showed no information about displacement to be available to the cognitive system under the conditions where pointing was affected (Bridgeman and Stark, 1979).

A more rigorous way to separate cognitive and motor systems was to put a signal only into the motor system in one condition and only into the cognitive system in another. We know that induced motion affects the cognitive system, because we experience the effect and subjects can make verbal judgments of it. But the above experiments implied that the information used for pointing might come from sources unavailable to perception. We inserted a signal selectivity into the cognitive system with stroboscopic induced motion (Bridgeman, Kirch, and Sperling, 1981). A surrounding frame was displaced, creating the illusion that the target had jumped, although it remained fixed relative to the subject. Target and frame were then extinguished, and the subject pointed open-loop to the last position of the target. Trials where the target had seemed to be on the left were compared with trials where it had seemed to be on the right. Pointing was not

significantly different in the two kinds of trials, showing that the induced-motion illusion did not affect pointing.

Information was inserted selectively into the motor system by asking each subject to adjust a real motion of the target, jumped in phase with the frame, until the target was stationary. Thus the cognitive system specified a stable target. Nevertheless, subjects pointed in significantly different directions when the target was extinguished in the left or the right positions, showing that the difference in real target positions was still available to the motor system. The visual system must have picked up the target displacement, but not reported it to the cognitive system, or the cognitive system could have ascribed the visually specified displacement to an artifact of frame movement. Thus a double dissociation occurred: in one condition the target displacement affected only the cognitive system, and in the other it affected only the motor behavior.

Dissociation of cognitive and motor function has also been demonstrated for the oculomotor system by creating conditions in which cognitive and motor systems receive opposite signals at the same time. Again the experiment involved stroboscopic-induced motion; a target jumped in the same direction as a frame, but not far enough to cancel the induced motion. The spot still appeared to jump in the direction opposite the frame, while it actually jumped in the same direction. Saccadic eye movements followed the veridical direction even though subjects perceived stroboscopic motion in the opposite direction (Wong and Mack, 1981). If a delay in responding was required, however, eye movements followed the perceptual illusion, implying that the motor system has no memory and must rely on information from the cognitive system under these conditions.

All of these experiments involve motion or displacement, leaving open the possibility that the dissociations are associated in some way with motion systems rather than with representation of visual space per se. A new series of experiments in my laboratory, however, has demonstrated dissociations of cognitive and motor function without any motion of the eye or the stimuli at any time. The dissociation is based on the Roelofs effect (Roelofs, 1935), a tendency to misperceive target position, in the presence of a surrounding frame presented asymmetrically, in the direction opposite the offset of the frame. The effect is similar to a stroboscopic induced motion in which only the final positions of the target and frame are presented (Bridgeman and Klassen, 1983).

# METHOD

## Subjects

The subjects were nine undergraduate volunteers and the author. Six of the subjects were naive with respect to the purposes of the experiment; the others assisted with the experiments, as well as serving as subjects.

## Apparatus

Subjects sat with stabilized heads before a hemicylindrical screen that provided a clear field of view 180° wide x 50° high. A rectangular frame 21° wide x 8.5° high x 1° in width was projected, via a galvanic mirror, either centered on the subject's midline, 5° left, or 5° right of center. Inside the frame, an "x" 0.35° in diameter could be projected via a second galvanic mirror in one of

five positions, 2° apart, with the middle "x" on the subject's midline (Fig. 1). A pointer with its axis attached to a potentiometer mounted near the center of curvature of the screen and its tip near the screen gave a voltage proportional to the tip's position, with a simple analog circuit. The voltage was fed into an A/D converter of a laboratory computer that controlled trial presentation and data collection. Perceived target position was recorded from a detachable computer keyboard placed in front of the subject. All keys except the five keys corresponding to the five target positions were masked off.

## PROCEDURE

### Training

Subjects were first shown the five possible positions of the target in sequence on an otherwise blank screen. Then they saw targets exposed for 1 sec and estimated their positions with the five response keys ("judging trials"), until they were correct in five consecutive trials. Next, they were trained on pointing, with the same stimuli ("pointing trials"), until they spontaneously returned the pointer to its rightmost position (as initially instructed) for five consecutive trials. In both conditions, subjects were instructed to wait until the offset of the stimulus before responding. Presentation of the target alone forced the subjects to use an egocentric judgment, and the long display time reduced the possibility of target onset eliciting a spurious motion signal that might affect responses.

### No Delay Condition

The 30 types of judging and pointing trials were mixed in a pseudorandom order. Each trial type was repeated 5 times, for a total of 150 trials/block. Trial order was restricted so that pointing trials and judging trials with the same target and frame positions would alternate in the series. At stimulus offset, subjects heard a short "beep" tone to indicate a judging trial or a longer "squawk" tone to indicate a pointing trial. There was a rest period after each 50 trials.

Trials were collated by the computer and a separate two-way ANOVA was run for each response type (assessing target main effect, frame main effect, and interaction).

### Delay Condition

Procedures were the same except that a 4-sec interval was interposed between stimulus offset and the tone that indicated the type of response.

# RESULTS

## No Delay Condition

For all subjects, there was a significant main effect of target position in both trial types and a significant main effect of frame position for judging trials. Thus, all subjects showed a Roelofs effect (Fig. 2).

The main effect of frame position in pointing trials showed a sharp division of the subjects into two groups: 5 of the 10 subjects showed a highly significant Roelofs effect ($p < 0.005$), while the other 5 showed no sign of an effect ($p > 0.18$). Thus, responses to pointing and judging trials were qualitatively different for half of the subjects, showing a Roelofs effect only for judging.

Four of the five subjects who showed a Roelofs effect in pointing were females. Thus, a sex effect is possible in this condition, with females more likely to code the target position in a symbolic form. The number of subjects, however, is too small to draw firm conclusions on this issue.

## Delay Condition

With a 4-sec delay interposed between display offset and tone, 9 of the 10 subjects showed a significant Roelofs effect for the judging task ($p < 0.01$) and 8 of the 10 showed a significant effect for the pointing task. One of the two remaining subjects showed no significant effect of frame position for either task. The other subject whose pointing behavior still showed no effect of the frame (Fig. 3) was retested with an 8-sec delay between display offset and tone. A Roelofs effect was found for both pointing and judging trials ($p < 0.001$) (Fig. 4).

In summary, interposting a long enough delay before the response cue forces all subjects to use pointing information that is vulnerable to bias from the frame position, even though half of the subjects were not vulnerable to this bias when responding immediately.

# DISCUSSION

These experiments show that perception of a Roelofs effect is robust, being seen by nearly all subjects under all delays. The Roelofs effect in visually guided behavior, though, depends much more strongly on the subjects and conditions. Half of the subject showed an effect of a surrounding frame on pointing behavior. The remainder showed the effect only when a long enough delay was interposed between target presentation and response.

The appearance of the Roelofs effect with a delay between stimulus and motor response is reminiscent of the results of Wong and Mack (1981): saccadic eye movements followed a veridical motion with a short delay, but followed a perceived motion in the opposite direction after a longer delay. If eye movements and visually guided behavior of the arm were controlled by a single motor-oriented internal map of the visual world, then we would expect the effects of delay to

influence eye and arm similarly, and the Wong and Mack results and our results could be explained in the same way.

There is now some evidence that oculomotor and skeletal motor systems do indeed share one map of visual space (Nemire and Bridgeman, 1987). Normally, eye and hand behavior are not correlated (Prablanc et al., 1979), in our interpretation because eye and hand motor systems read their information from the same visual map through separate, independent noise sources. To show the identity of visual information driving these two systems, we disturbed the normally veridical mapping process by having subjects make repeated saccades in darkness. This resulted in saccade undershoot, but equally great undershoot of manual pointing.

Our conclusion is that the normal human possesses two maps of visual space. One of them holds information used in perception: if subjects are asked what they see, the information in this "cognitive" map is accessed. The other map drives visually guided behavior, for both eye and arm. The "motor" map is not subject to illusions such as induced motion and the Roelofs effect. In this sense it is more robust, but as a result it is less sensitive to small motions or fine-grained spatial relationships. It also has no memory, being concerned only with the here-and-now corre- spondence between visual information and motor behavior. If a subject must make motor responses to stimuli no longer present, this system must take its spatial information from the cog- nitive representation, and brings any cognitively based illusions along with it.

An alternative explanation of the results has been suggested (Ian Howard, personal communi- cation, Sept. 2, 1987); presentation of an off-center frame might bias the subject's subjective straight-ahead in the same direction as the frame's offset. Judging of point position would then be biased in the opposite direction because the subject bases his or her judgments on an offset straight ahead direction. Pointing, however, would remain the same because the subject has not in fact moved, and arm position must be egocentric. This alternative can be tested empirically by having subjects point to the center of the apparatus when the frame is presented in center, left, or right position. Preliminary data from three subjects indicate that frame position has no effect on pointing straight ahead.

Finally, we can apply this conception of two maps of visual space to design of spatial displays. Any display where perception is the primary goal, such as displays of the status of instruments, is subject to induced-motion illusions, Roelofs effects, and other cognitive biases. The designer can take advantage of these effects in designing such displays, but must beware that they do not distort the data displayed.

Displays which guide real-time behavior, on the other hand, are not subject to such illusions. The designer need not worry, for instance, about background motions affecting visually guided behavior toward a target (Bridgeman, Kirch, and Sperling, 1981). But information must be avail- able continuously, for the internal map guiding these behaviors has no significant memory.

# REFERENCES

Bridgeman, B.; Hendry, D.; and Stark, L. (1975): Failure to detect displacement of the visual world during saccadic eye movements. Vision Res., 15, 719-722.

Bridgeman, B.; Kirch, M.; and Sperling, A. (1981): Segregation of cognitive and motor aspects of visual function using induced motion. Percept. Psychophys., 29, 336-342.

Bridgeman, B.; and Klassen, H. (1983): On the origin of stroboscopic induced motion. Percept. Psychophys., 34, 149-154.

Bridgeman, B.; Lewis, S.; Heit, G.; and Nagle, M. (1979): Relation between cognitive and motor-oriented systems of visual position perception. J. Exp. Psychol: Human Percept. Perform., 5, 692-700.

Bridgeman, B.; and Stark, L. (1979): Omnidirectional increase in threshold for image shifts during saccadic eye movements. Percept. Psychophys., 25, 241-243.

Brune, F.; and Lücking, C. H. (1969): Oculomotorik, Bewegungswahrnehmung und Raumkonstanz der Sehdinge. Der Nervenarzt, 40, 692-700.

Ditchburn, R. (1955): Eye-movements in relation to retinal action. Optica Acta, 1, 171-176.

Festinger, L.; and Canon, L. K. (1965): Information about spatial location based on knowledge about efference. Psychol. Rev., 72, 373-384.

Hallett, P. E.; and Lightstone, A. D. (1976): Saccadic eye movements towards stimuli triggered during prior saccades. Vision Res., 16, 99-106.

Mack, A. (1970): An investigation of the relationship between eye and retinal image movement in the perception of movement. Percept. Psychophys., 8, 291-298.

Nemire, K.; and Bridgeman, B. (1987): Oculomore and skeletal motor systems share one map of visual space. Vision Res., 27, 393-400.

Prablanc, C.; Echallier, J. F.; Komilis, E.; and Jeannerod, M. (1979): Optimal response of eye and hand motor systems in pointing. Biolog. Cybern., 35, 113-124.

Roelofs, C. (1935): Optische localisation. Archives für Augenheilkunde, 109, 395-415.

Van Essen, D. C.; Newsome, W. T.; and Bixby, J. L. (1982): The Pattern of interhemispheric connections and its relationship to extrastriate visual areas in the macaque monkey. J. Neurosci., 2, 265-283.

Wallach, H.; and Lewis, C. (1965): The effect of abnormal displacement of the retinal image during eye movements. Percept Psychol., 1, 25-29.

Wong, E.; and Mack, A. (1981): Saccadic programming and perceived location. Acta Psychol., 48, 123-131.
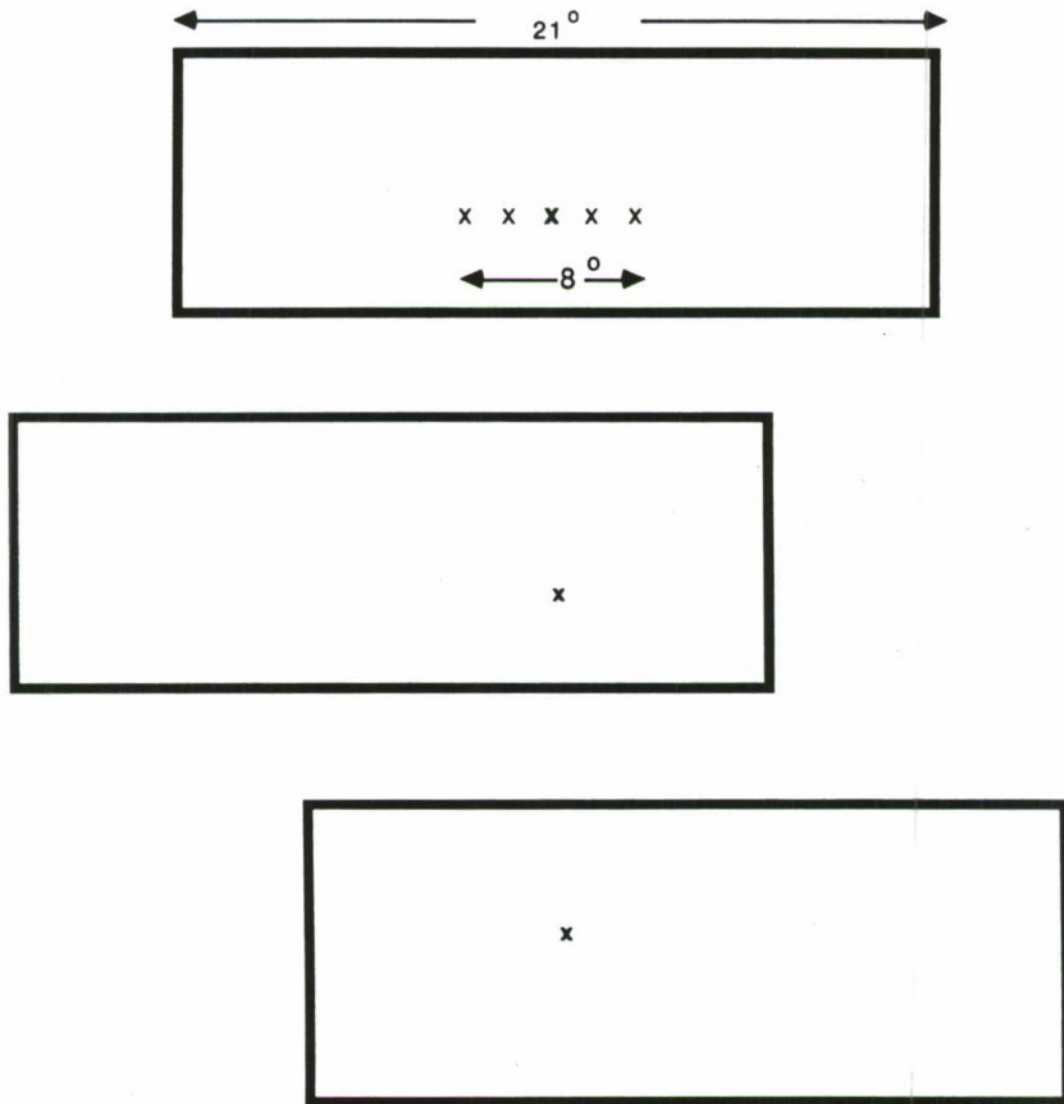
Figure 1.– Stimulus array used in pointing/judging experiments. The frame could be centered (top), biased 5° left (middle), or biased 5° right (bottom). A target appeared in one of the five positions indicated in the top frame. Other frames show the position of the center target.
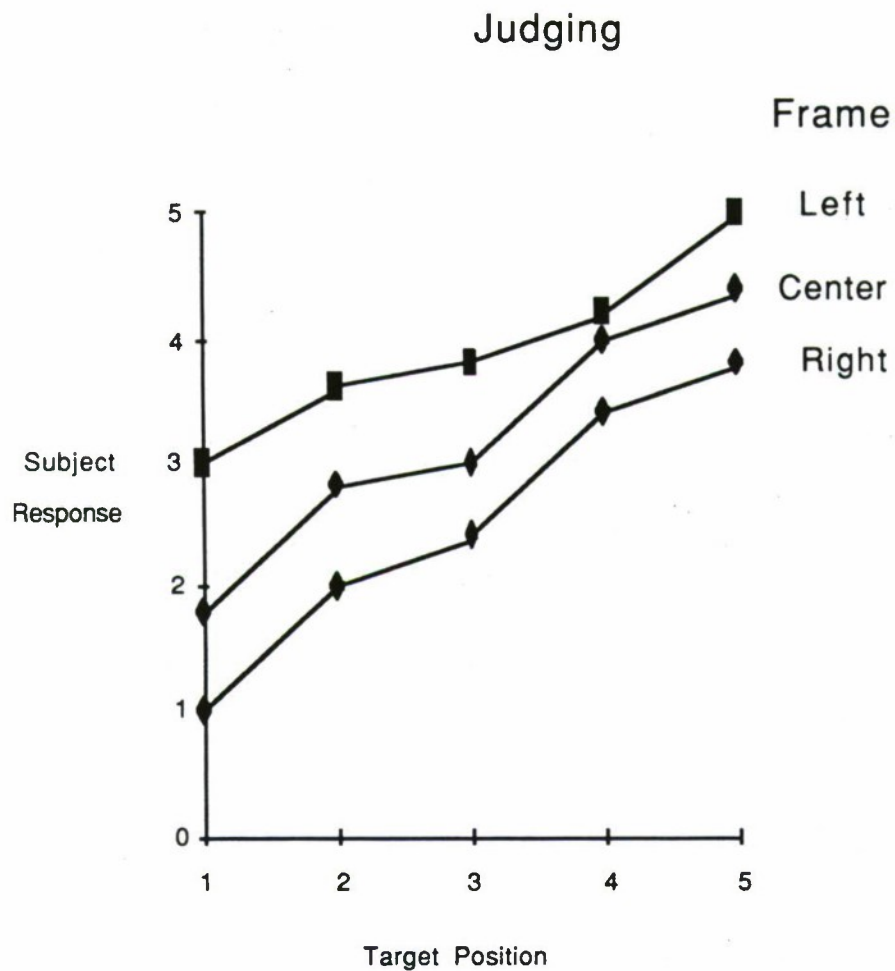
Figure 2.– Judging and pointing behavior immediately after stimulus offset. a) Judging target position with a five-alternative, forced-choice procedure. The separation of three curves corresponding to the three frame positions is due to the Roelofs effect.
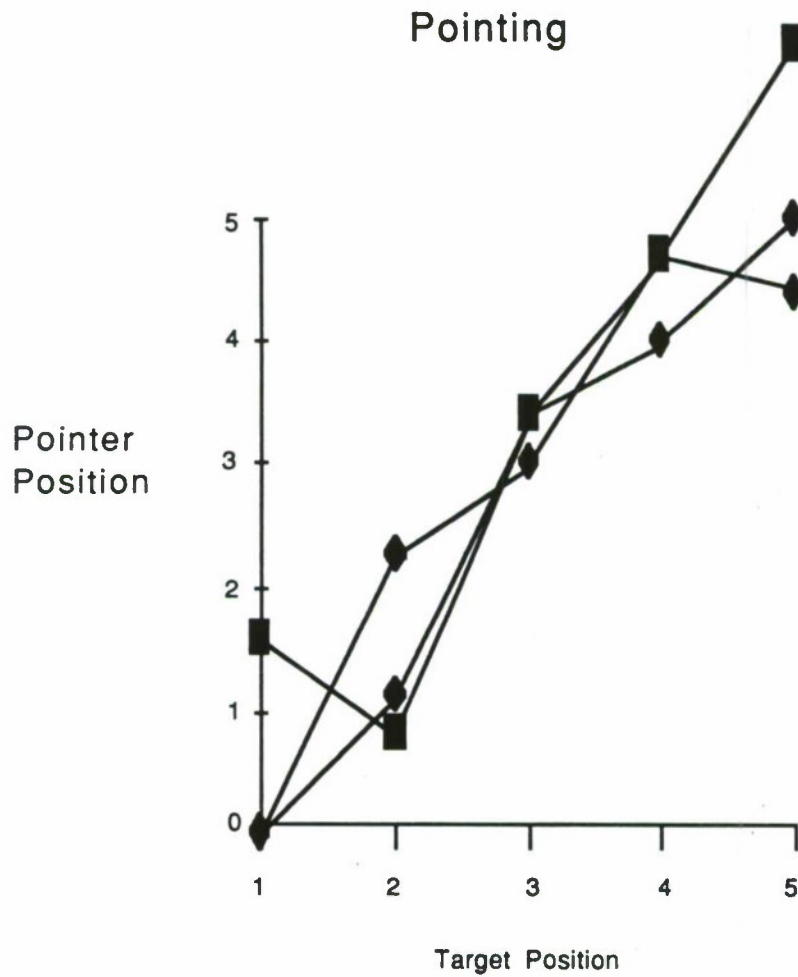
Figure 2.– Concluded. b) Pointing to targets under the same perceptual conditions, in trials intermingled with the judging trials. Overlap of the three curves indicates lack of influence of frame position on pointing behavior. Data are from one subject.

# Judging
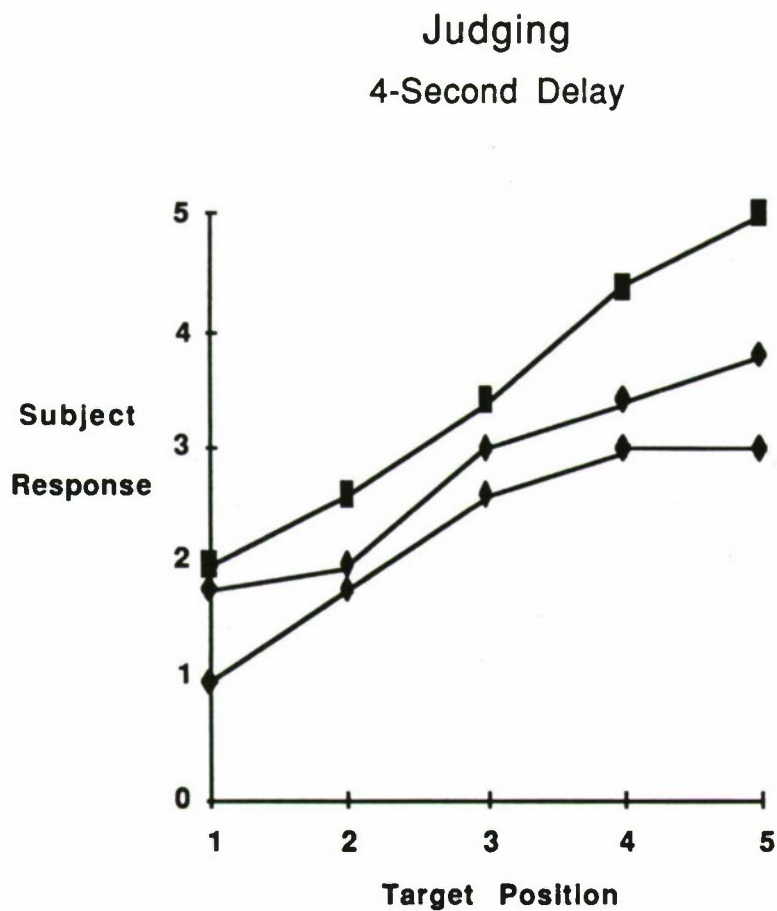
## 4-Second Delay



Figure 3.– Judging and pointing after a 4-sec delay. In this subject, no Roelofs effect is evident for pointing; the other subjects showed an effect at this delay.

Pointing

4-Second Delay
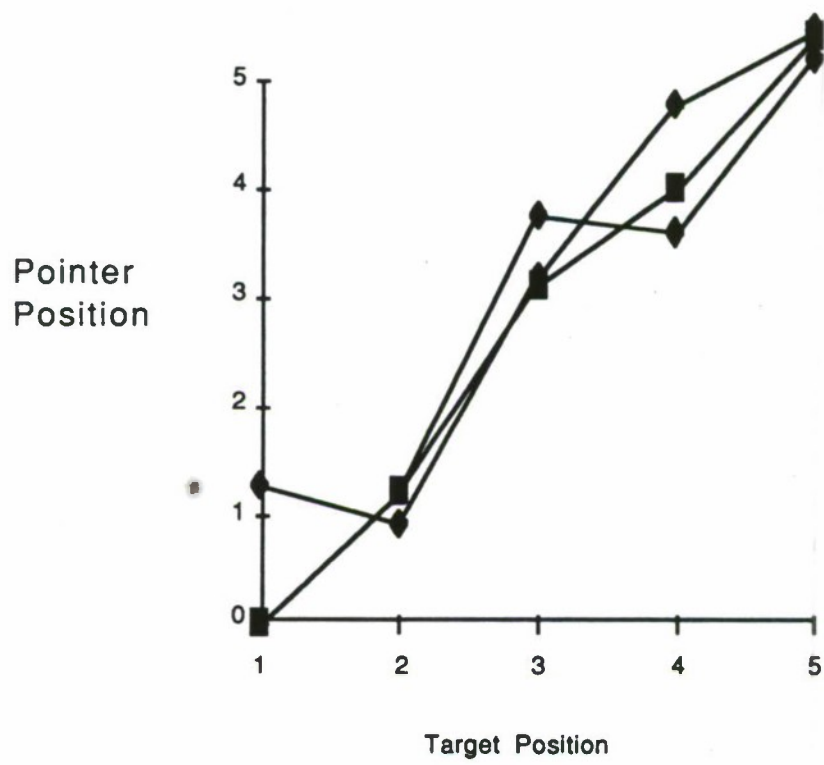
Figure 3.– Concluded.
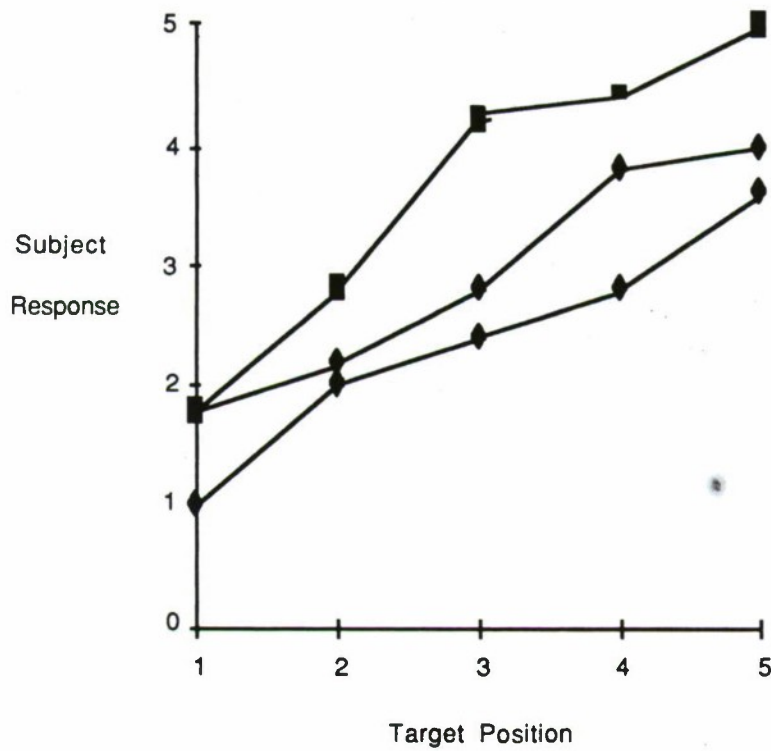
Figure 4.– Judging and pointing after an 8-sec delay. A Roelofs effect for pointing has appeared.
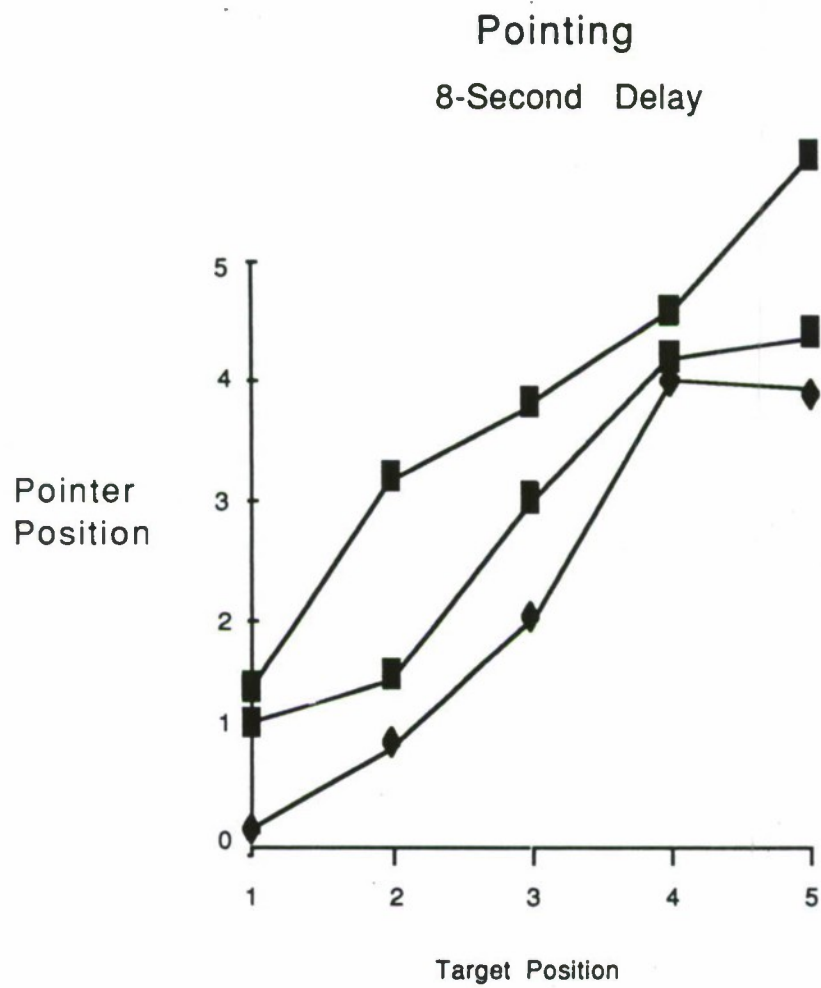
Figure 4.– Concluded.

# PICTURE PERCEPTION:  PERSPECTIVE CUES

# THE EFFECTS OF VIEWPOINT ON THE VIRTUAL SPACE OF PICTURES

H. A. Sedgwick
Schnurmacher Institute for Vision Research
S.U.N.Y. College of Optometry, New York, New York

## 1. INTRODUCTION

Pictures are made for many different purposes (Hagen, 1986; Hochberg, 1979). This discussion is about pictorial displays whose primary purpose is to convey accurate information about the three-dimensional spatial layout of an environment. We should like to understand how, and how well, pictures can convey such information. I am going to approach this broad question through another question that seems much narrower. We shall find, however, that if we could answer the narrow question, we should have made a good start on answering the broader question as well.

Every pictorial display that presents a precise perspective view of some three-dimensional scene has a single geometrically correct viewpoint.[1] In most viewing situations, however, the observer is not constrained to place his or her eye precisely at this correct viewpoint; indeed the observer generally has no explicit knowledge of the location of this viewpoint.[2] My "narrow" question is: "What effect does viewing a picture from the wrong location have on the virtual space represented by that picture?"

This question is in itself of theoretical as well as practical importance. It has received considerable attention, but its answer is still far from being clear. The research literature is fragmentary and conflicting. I believe that a more vigorously applied theoretical analysis can clarify the issues and can help in evaluating the existing literature.

My theoretical analysis follows the approach developed by J. J. Gibson (1947, 1950, 1954, 1960, 1961, 1971, 1979). I shall be referring frequently to the *optic array*, which is Gibson's term for the structured array of light reflected to a point of observation by the surfaces of the environment. I shall also be relying on Gibson's concept of *available visual information*. Information is said to be available in the optic array when some projective structure in the optic array mathematically specifies, with appropriate constraints, some structure in the environment. The optic array typically contains multiple, redundant sources of information for the spatial layout of the environment.

The theoretically determined availability of visual information of course does not guarantee that such information will be used by a human observer. The extent to which any such information actually influences perception is a separate question that must be addressed empirically. The contention of Gibson's approach is simply that we are not in a proper position to formulate or interpret empirical investigations of human visual perception until we understand the underlying available information on which any successful perception must be based.

This discussion will concentrate on theoretical analysis. At several points, however, I shall briefly indicate how well this analysis accords with the empirical work that has been done on

human pictorial perception. More detailed reviews of this subject are offered elsewhere (Cutting, 1986a; Farber and Rosinski, 1978; Hagen, 1974; Kubovy, 1986; Rogers, 1985; Rosinski and Farber, 1980).

To simplify the discussion I am going to consider separately the effects of deviating from the correct viewpoint in each of three orthogonal directions: deviations perpendicular to the picture plane (that is, being too close or too far from the picture), lateral deviations parallel to the picture plane, and vertical deviations parallel to the picture plane. Any possible viewing position can then be interpreted as some combination of these three deviations.

## 2. THEORETICAL ANALYSIS

### 2.1 Viewing from Too Close or Too Far

What is the theoretical effect of viewing a pictorial display from too close or too far[3]? As we approach or withdraw from the picture, its projection in the optic array expands or contracts around the center of the picture, which is the point at which a perpendicular from the viewpoint pierces the picture plane. If we let $z$ be the correct distance from the picture and $z'$ be our actual distance, and let $A$ and $A'$ be the angular separations from the center at these two distances, respectively, of some other point on the picture, then

$$\tan A / \tan A' = z'/z = m$$

where $m$ is a constant. Thus the optic array projection of the picture is magnified or minified by $1/m$, where $m$ measures how close or how far we are, relative to the correct distance.[4]

What, in theory, is the effect on the virtual space of the picture of magnifying or minifying its projection in the optic array? We can begin to answer this question by looking at the available visual information that is present in the *perspective structure of the optic array*, by which I mean the vanishing points of straight edges in the environment and the vanishing lines of planar surfaces.[5]

Let us imagine a picture of a flat, endless ground plane covered with a regular texture represented by a grid of lines. The horizon, or vanishing line, of the ground, will be located at eye level on the picture plane. If our point of observation is located at a height $h$ above the ground, then the distance $d$ along the ground to any particular grid line parallel to the picture plane is given by the simple expression

$$d = h(1/\tan G)$$

where $G$ is the optic array angle subtended between the horizon of the ground plane and the grid line.

We can now combine these two expressions to derive the theoretical effect of magnification or minification. If we let $d'$ be the geometrically specified distance of the grid line when the picture is seen from the incorrect viewpoint and let $G'$ be the new optic array angle corresponding to $G$, then

$$d' = h(1/\tan G')$$

substituting for G',

$$d' = h(m/\tan G)$$

and substituting again,

$$d' = md$$

Next, if we let s be the specified separation in depth between any two successive grid lines, at distances d1 and d2, when the picture is seen from the correct viewpoint and let s', d1', and d2' be the specified separation and distances when seen from the incorrect viewpoint, then

$$
\begin{aligned}
s' &= d2' - d1' \\
&= md2 - md1 \\
&= m(d2 - d1) \\
&= ms
\end{aligned}
$$

Thus as we approach the picture, the geometrically specified depths in the picture are compressed proportionally to the closeness of our approach and as we move away from the picture, depths are expanded proportionally (fig. 1).[6]

Consider now what happens to frontal plane dimensions. The tangent of the angle F subtended by a width w that is parallel to the picture plane is inversely proportional to its distance from the point of observation (assuming for simplicity that the width is measured from the center of the picture)

$$w = d \tan F$$

As we approach the picture, the specified distance of w decreases, but its optic array angle F increases in the same proportion, so that w remains constant (fig. 2)

$$w' = d' \tan F' = (md)(\tan F/m) = d \tan F = w$$

The depth of the pictured scene is thus compressed relative to its frontal dimensions. Shapes that are not in the frontal plane are distorted. The square grid covering the ground plane, for example, becomes a grid of rectangles whose depth to width ratio is m (fig. 3).

We may note here that all distances specified in the virtual space of the picture depend on h, the height of the viewpoint above the ground plane, which thus provides a scale factor for all distances, as well as sizes, in the picture. Because h itself is not geometrically specified in the picture, its value may be indeterminate.[7] This indeterminancy of h puts in doubt the appropriateness of comparing absolute distances or sizes across different pictures or across different views of the same picture. The ratio, however, of depth to width, s/w or s'/w', does not depend on h; thus, geometrically specified compression of shape by the factor m is an invariant effect of too close viewing.

Geometrically specified angles and orientations in the pictured scene are also changed by approaching the picture. This result follows directly from the compression that occurs, but it is instructive to derive the result in a different way.

Every set of parallel lines in the pictured scene has a vanishing point on the picture plane (lines parallel to the picture plane have their vanishing points at infinity on the picture plane). The three-dimensional orientation of a set of parallel lines is equal to the orientation of a line from the point of observation to their vanishing point. This very simple optic array relation specifies the pictured orientation of any edge once its vanishing point is known (Hay, 1974; Sedgwick, 1980).

Edges perpendicular to the picture plane have their vanishing point at the center of the picture. As we approach the picture, every vanishing point except for the one at the center of the picture increases its optic array separation from the central vanishing point. Thus the specified orientations of all nonperpendicular edges move closer to being parallel to the picture plane. For example, a square ground plane grid oriented at 45° to the picture plane becomes a grid of squashed diamonds (fig. 4).

If we let E be the angle, measured relative to the straight-ahead, that a vanishing point subtends at the correct viewpoint, and let E' be the angle that it subtends when the viewpoint is too close or too far, then the distortion D in the specified orientation of edges having that vanishing point is given by E minus E'.[8] The relation between E and E' is the same as for any other optic array angles measured from the center of the picture, namely

$$\tan E / \tan E' = m$$

Calculating D as a function of E for several values of m, we obtain a family of curves showing no distortion for orientations perpendicular (0°) or parallel (-90° or 90°) to the picture plane, with maximum distortion at intermediate values (fig. 5). For example, for m equal to either 2 or 0.5, the maximum distortion approaches 20°.

A similar analysis can be made for the orientations of planar surfaces. The angle subtended between the vanishing line of a slanted surface and the vanishing line of the ground plane is equal to the three-dimensional angle between the depicted surface and the ground (Sedgwick, 1980). As we approach the picture plane, geometrically specified surface orientations are distorted in just the same way as are edge orientations.

Perceptually, effects qualitatively similar to those predicted theoretically here can be seen by a careful observer moving closer or farther from a picture containing strong linear perspective. If the perspective information in the picture is weaker, the distortions may be much harder to see. Most empirical investigations, but not all, have found such distortions in human picture perception, although not always at the magnitude predicted.[9] I shall say a bit more about the reasons for the discrepancies between investigations later.

## 2.2  Viewing from the Side

Let us now consider what happens when we view a pictorial display from the side.[10] It is easy to see that when the viewpoint is displaced laterally, maintaining the same distance from the picture plane, the horizon of the ground and all of the grid lines parallel to the picture plane simply

slide along themselves in the optic array. Thus the angular separation of each of these grid lines from the ground horizon remains unchanged. Consequently, the geometrically specified distance of each of these grid lines, relative to the height of the viewpoint, also is unchanged (fig. 6).

As the viewpoint slides to the right, for example, each point in the geometrically specified virtual space of the picture slides to the left, with its projected point on the surface of the picture acting as a stationary fulcrum. This lateral shift in virtual space is thus directly proportional to, but opposite in sign from, the amount of the viewpoint's displacement; it is also directly proportional to the distance of the point from the picture plane, and is inversely proportional to the viewpoint's distance from the picture plane (fig. 7). The overall effect of this viewpoint displacement is to produce a lateral shear in the geometrically specified virtual space of the picture (fig. 8). Frontal plane dimensions and orientations are unchanged, but shapes and orientations extending in depth are all distorted.

We can readily determine the specified shifts in the orientations of pictured edges and surfaces by again making use of the perspective structure of the picture. Let us consider, as an example, the orientations of horizontal edges, whose vanishing points lie on the horizon of the ground plane. As the viewpoint shifts laterally, its angular relation to each of these vanishing points changes. We shall let E again be the angle, measured relative to the straight-ahead, that the vanishing point makes with the correct viewpoint, and let E' be the angle that it makes after the vanishing point has shifted laterally. We can express this lateral shift as the ratio, k, between the amount, r, of the shift, and the distance, z, of the viewpoint from the picture plane. It is easy to see that (fig. 9)

$$\tan E' = \tan E + k$$

If we express the position of the shifted viewpoint in terms of its angular deviation, V, from the correct viewpoint, then

$$\tan V = k$$

so that

$$\tan E' = \tan E + \tan V$$

We can use this relation to determine the specified distortion of orientation, E' minus E, as a function of the correct orientation E, for a variety of angular shifts V of the viewpoint (fig. 10). The resulting family of curves shows that the specified distortions in orientation can be very large, approaching 180° as V approaches ±90°, which is parallel to the picture plane, and that the orientation E at which the distortion is maximal increases as V increases.

We may note that the same distortions in orientation would also be specified for vertical planes in the virtual space of the picture when the viewpoint is displaced laterally.

Perceptually, again, a careful observer comparing the appearance of a picture seen from one side or the other can notice differences in apparent orientation if the picture contains sufficient perspective information. Some empirical investigations have also found results that are qualitatively similar to those derived here, although others have not.[11] Again, I shall refer back to these discrepancies a little later.

## 2.3 Viewing from Too High or Too Low

Let us now briefly consider what happens when the viewpoint is too high or too low. This is again a displacement parallel to the picture plane, so the geometrically specified distortions in the virtual space of the picture are identical in form to those produced by lateral shifts, except that here the virtual space is sheared vertically instead of laterally.

Thus, for example, if we consider a plane in virtual space that is rotated around a horizontal axis so that it makes an angle E with the ground, its specified slant E', when seen from an incorrect viewpoint having a vertical angular deviation V, is given by the same relation

$$\tan E' = \tan E + \tan V$$

Notice that if we are considering the ground plane itself, then E = 0, so that E' = V. That is, if we must look down by a certain angle to see the pictured horizon, then the ground plane is specified as slanting down by that same angle.

## 3. THEORETICAL COMPLICATIONS

So far we have seen how we can use the perspective structure of the optic array to determine the geometrically specified sizes, distances, and orientations of surfaces and edges in the virtual space of a picture. We have also seen how this visual information, when it is present, specifies distortions in the pictured layout when we observe the picture from the wrong viewpoint. Unfortunately for our ease of understanding, there are theoretical complications that are not taken into account by this straightforward analysis. We need to consider some of these complications now.

### 3.1 Resolving Multiple Sources of Visual Information

In a normally complex pictorial display, there are available other sources of visual information for spatial layout besides those arising from the perspective structure of the picture. How these multiple sources of information, which are normally partially redundant and partially complementary, may be combined into a single perceptual interpretation is a difficult and as yet unsettled question.[12] The difficulty is increased when the picture is observed from the wrong viewpoint because these different sources of information do not all predict the same distortions; nor is it always easy to tell what they do predict.

As an example, consider some of the information arising from surface texture (Gibson, 1950; Sedgwick, 1983, 1986). If several edges are resting on a surface that is uniformly textured, then the relative lengths of the edges are specified by the relative amounts of texture that they cover; likewise, the relative distances between the edges are specified by the relative amounts of texture between them. This texture scale information is as valid for edges that extend into depth as for those in the frontal plane; it thus serves to specify the shapes and the relative sizes and distances of objects resting on a common textured surface such as the ground plane.

It is easy to see that all such texture scale information is completely invariant over changes in viewpoint because such changes do nothing to alter the depicted amounts of texture between or under the objects in the picture. If, for example, we approach the picture of a square object resting on the textured ground, the specified object remains square because each of its edges continues to cover an equal amount of texture. On the other hand, according to the analysis based on perspective structure, the specified object is compressed into a rectangle whose width is greater than its depth.

This apparent contradiction between the distortions predicted by these two sources of visual information can be resolved, but only in a way that further complicates our analysis. I mentioned earlier that any visual information entails constraints on the environment; if these constraints are violated, then the information is no longer valid. In the case of texture scale information, an essential constraint is that the texture's distribution across a surface be at least statistically uniform. Yet, in the example that we are considering now, when we come too close to the picture, perspective analysis specifies that the texture of the ground is itself compressed in the depth dimension. Thus the uniform distribution constraint is violated and texture scale information is no longer valid.

A visual system might do any of a number of things when faced with this situation. It might simply reject texture scale information as being invalid. It might go ahead and use texture scale information anyway. It might recognize that the viewpoint is incorrect. It might abandon the attempt to find a consistent virtual space for the picture. It might adopt a modified version of texture scale information using compressed texture. It might do something intermediate between some of these options. Analysis only indicates the possibilities without specifying which one will be adopted by any particular visual system.

A number of other sources of visual information, such as right-angle constraints (Perkins, 1972, 1976) and orientation-distribution constraints (Witkin, 1980), present similar difficulties when the viewpoint is incorrect, but there is not space to consider these additional difficulties here. Careful analysis of the interactions between these different sources of information should give us a basis for manipulating the information content of pictures so as to better determine the perceptual effects they produce.

### 3.2 Constancy and the Dual Nature of Pictures

A second set of theoretical complications arises from what has often been referred to as the "dual nature" of pictures (Gibson, 1954; Haber, 1979, 1980a, 1980b; Hagen, 1974, 1986; Hochberg, 1962, 1979; Pirenne, 1970). In addition to being a representation of a spatial layout existing in a three-dimensional virtual space that lies beyond the plane of the picture, a pictorial display is also a real object consisting of markings of some sort, usually on a flat surface. Normally, visual information for the flat surface of the picture is made available by binocular stereopsis, by motion parallax, by the oculomotor adjustments of convergence and accommodation, by the frame of the picture, and by the surface texture of the picture.

To perceive pictures, a perceptual system must be able, to some extent, to differentiate its response to the picture's virtual layout from its response to the real layout of the picture's surface. The human visual system seems able to make this differentiation, but not without some interaction, or "cross talk," between its responses to these two classes of information.

We can get some understanding of one effect of the picture surface by examining the relation between the picture plane and the optic array. If x measures a separation in the picture plane from the center of the picture, which we have already defined as the point where a perpendicular from the viewpoint pierces the picture plane, and A measures the optic array angle subtended by this separation, then x is related to A by the relation

$$x = z \tan A$$

where z is the distance from the viewpoint to the picture plane. Near the center of a picture there is a close congruence between the optic array projection and the flat picture plane projection. This is because the tangent function is nearly linear for small angles. For larger angles, however, the tangent function becomes highly nonlinear, and consequently the optic array projection and the picture plane projection become strongly noncongruent.

Perceptually, the cross talk between the picture surface and the virtual space of the picture, as specified in the optic array, becomes most noticeable when the picture plane projection and the optic array projection are noncongruent. Toward the edges of wide-angle pictorial displays, for example, the projections on the picture plane and in the optic array are still geometrically correct, but objects in the virtual space of the picture often appear to be distorted (Pirenne, 1970, 1975; Kubovy, 1986).[13] It seems that the noncongruent shape on the surface of the picture takes on a perceptual salience that interacts with the virtual space of the picture.

A similar noncongruence between the picture plane and the optic array is produced when the viewpoint is displaced laterally or vertically from the correct viewpoint. Again, the noncongruent shape on the surface of the picture may interact perceptually with the virtual space of the picture, but here its effect would be to diminish the distortion that is specified in the optic array. This would result in some degree of "constancy" in the virtual space of the picture in the sense that the virtual layout would not be as distorted as the optic array information would predict.

These effects of the picture's surface on the perceived virtual space of the picture could be eliminated, in principle, by removing the visual information for the picture's surface. Using a monocular display, restricting head movements relative to it, hiding the frame of the display, and so on, would all contribute to this result (Ames, 1925; Enright, 1987; Schlosberg, 1941; P. C. Smith and O. W. Smith, 1961).

### 3.3 The Hypothesis of Pictorial Compensation

Finally, many theorists have suggested that when information for the picture surface is available, the human visual system may be able to compensate for being at the wrong viewpoint and so avoid distortions in the virtual space of the picture (Cutting, 1987; Farber and Rosinski, 1978; Hagen, 1974, 1976a, 1976b; Kubovy, 1986; Perkins, 1973, 1980; Pirenne, 1970; Rosinski, 1976; Rosinski and Farber, 1980; Rosinski, Mulholland, Degelman, and Farber, 1980; Wallach and Marshall, 1986). This compensation process would operate by either detecting or assuming a "correct" position of the viewpoint. The optic array information would then be adjusted to determine the virtual layout as it would be seen from this correct viewpoint.

Although a number of experiments have been offered in support of this view, it seems to me that, on balance, the compensation hypothesis is neither necessary nor sufficient to account for

the bulk of the empirical results. It is not necessary because, as we have just seen, however sketchily, there are other explanations available for some of the disparities that exist between the distortions predicted by perspective structure and those actually found. Moreover, these other explanations are more parsimonious, in that they are derived from the analysis of general perceptual processes without having to postulate special processes that exist solely for perceiving pictures from the wrong viewpoint. The compensation hypothesis is not sufficient because it does not account for the considerable number of experimental results that find distortions in virtual space even when there is information available for the surface of the picture (Bengston, et al., 1980; Goldstein, 1979, 1987; Wallach, 1976, 1985). Finally, it seems to me that a careful reading of several of the key experiments offered in favor of the compensation hypothesis casts some doubt on the firmness of their conclusions.[14]

## 4. CONCLUSION

As a conclusion to this brief discussion, I would suggest that picture perception is not best approached as a unitary, indivisible process. Rather, it is a complex process depending on multiple, partially redundant, interacting sources of visual information for both the real surface of the picture and the virtual space beyond. Each picture must be assessed for the particular information that it makes available. This, I would suggest, will determine how accurately the virtual space represented by the picture is seen, as well as how it is distorted when seen from the wrong viewpoint.

# NOTES

1.    For a camera image, this point is determined by the optics of the imaging system; for a display created by a draftsman or a computer, this point is determined by the relation between the center of projection and the projection plane (Carlbom and Paciorek, 1978; Sedgwick, 1980).

2.    A complex pictorial display generally does contain sufficient information, under certain constraints, to specify its own correct viewpoint. This issue is discussed by Green (1983), Jones and Hagen (1978), and Sedgwick (1980).

3.    A number of analyses of this problem have been offered. The first systematic analysis appears to come from La Gournerie (1859), whose work has been discussed more recently by Pirenne (1970, 1975), Kubovy (1986), and Cutting (1987). Other analyses, apparently independent of La Gournerie, have been given by Purdy (1960), Farber and Rosinski (1978), Lumsden (1980), and Rosinski and Farber (1980).

Obtaining an unambiguous three-dimensional interpretation of a pictorial display requires that some constraints be placed on the possible interpretations. In the above analyses, those referring to La Gournerie and that of Farber and Rosinski (1978) do not make these constraints explicit. The other analyses use explicit constraints derived from analyses of normally viewed pictures. Purdy (1960) bases his analysis on gradients of texture, Lumsden bases his on familiar size, and Rosinski and Farber base theirs on linear perspective. I offer two analyses here, one based on the ground plane and the other based on perspective structure, as suggested in Sedgwick (1980). All of these analyses converge on the same results.

A different analysis, reaching different results, has been offered recently by McGreevy and his colleagues (Ellis et al., 1985; McGreevy and Ellis, 1984, 1986; McGreevy, Ratzlaff, and Ellis, 1987). McGreevy's analysis proceeds by arbitrarily constraining all virtual distances from the picture plane to be unchanged by viewing position. This analysis has the weakness that it assumes a knowledge of these distances without indicating how they could be determined by an observer of the display, either when viewing from the wrong viewpoint or when viewing from the correct viewpoint. The question of how virtual layout could be determined here is made difficult because the constraint that is imposed leads to violations of all of the other constraints mentioned in the preceding paragraph.

Another kind of analysis, based on optimizing the match between a noisy registration of the projection and a noisy a priori internal model of the spatial layout has been offered recently by Grunwald and Ellis (1986). There is not room here to consider the interesting question of how such a model-based approach to spatial layout might be reconciled with the constraint-based approach taken in this paper.

4.    Approaching a picture is optically equivalent to viewing the pictured scene through a telephoto lens, and withdrawing from the picture is optically equivalent to viewing the scene through a wide-angle lens (Lumsden, 1980; Rosinski and Farber, 1980).

5.    Perspective structure is usually only implicit in the optic array. The available visual information that specifies this perspective structure is not discussed in this paper, but I have analyzed it

in detail elsewhere. Not all pictorial displays contain sufficient information to completely specify their perspective structure (Sedgwick, 1983, 1986, 1987a).

6.     There is an invariant associated with the optic array gradient projected from equally spaced grid lines parallel to the picture plane. If s is the separation in depth between any two successive grid lines, then

$$s = d2 - d1 = h(1/\tan G2 - 1/\tan G1)$$

Thus, for any two successive optic array angles G1 and G2 in this gradient

$$1/\tan G2 - 1/\tan G1 = k$$

where k is a constant. The presence of this invariant in the optic array specifies that the grid lines are equally spaced. It can be shown that this invariant is preserved when the picture is viewed from too close or too far.

7.     The value of h can be determined by assuming that the ground plane of the picture is coextensive with the ground plane of the real environment, but such an assumption may for some pictures be neither appropriate nor perceptually compelling.

8.     Throughout this paper, orientations are specified in environment-centered terms (i.e., relative to the fixed framework of the environment), rather than in viewer-centered terms (i.e., relative to the observer's line of regard). I have discussed this distinction and its significance at length elsewhere (Sedgwick, 1983; Sedgwick and Levy, 1985).

9.     Empirical evidence that is at least qualitatively consistent with the analysis presented here has been reported by Bartley (1951), Bartley and Adair (1959), Bengston et al. (1980), Farber (1972), Lumsden (1983), Purdy (1960), O. W. Smith (1958a, 1958b), O. W. Smith and Gruber (1958), and O. W. Smith, P. C. Smith, and Hubbard (1958). Anecdotal supporting observations are also reported by MacKavey (1980) and Pirenne (1970). On the other hand, Rosinski and Farber (1980) briefly report failing to find distortions when the frame of the display is visible, and Hagen and Elliott (1976) and Hagen and Jones (1978) report that adults' choice of the most "realistic looking" display was essentially independent of their actual viewing distance.

It is important to distinguish between the presence of measurable distortions in the perception of spatial layout and the detection of these distortions by the observer. Observers' perceptions may contain distortions of which the observers themselves are unaware. A number of researchers have suggested that observers are often not very sensitive to the presence of such distortions (Gombrich, 1972; Pirenne, 1970; Cutting, 1986a, 1986b).

10.     Systematic analysis of this problem is again offered by La Gournerie (1859), whose work has been put to use by Cutting (1987). More recent analyses are offered by Farber and Rosinski (1978) and Rosinski and Farber (1980), who explicitly base their second analysis (1980) on linear perspective constraints. I again offer two analyses, one based on the ground plane and the other, following Sedgwick (1980), based on perspective structure. All of these analyses agree in the distortions that they predict.

11.     Anecdotal reports of these distortions are common (Pirenne, 1970, 1975; Wallach, 1976, 1985). Experimental evidence that such distortions occur perceptually under some circumstances is offered by Goldstein (1979, 1987), Rosinski et al. (1980), Rosinski and Farber (1980), and Wallach and Marshall (1986), although all of these authors also report conditions under which the analytically predicted distortions do not occur. Cutting (1987) has analyzed some of the data of Goldstein (1987) in detail and has shown it to be in generally good accord with the theoretical predictions. Perkins (1973) finds some distortion from lateral viewing, but much less than this analysis would predict.

12.     An expert system that I have developed to study the interaction of multiple sources of visual information is described elsewhere (Sedgwick, 1987a, 1987b).

13.     This assumes that the perpendicular from the correct viewpoint pierces the picture plane somewhere near the center of the pictorial display, as it usually does.

14.     Kubovy (1986) is critical of many of the stimuli used by Hagen and Elliott (1976) and Hagen and Jones (1978) in their demonstration that adults at various distances from a picture do not choose the correct perspective as being most realistic. Perkins' (1973) demonstration of compensation for lateral viewing uses such minimal stimuli that the applicability of his results to more complex displays may reasonably be questioned. Hagen's (1976b) study, which claims to find evidence of compensation for lateral viewing in adults, has been criticized at length on logical grounds by Rogers (1985), who also failed to replicate Hagen's results. In the carefully controlled study of Rosinski et al. (1980) on the effects of frame visibility on perceived surface slant with lateral viewing, the interpretation of results is clouded by a confusion in the description of the experiment, and possibly in the experiment itself, about the frame of reference for their observers' judgments. Finally, Wallach and Marshall (1986, exp. 2) find evidence of compensation in pictorial shape perception from a lateral viewpoint, but their results, as they note, could be due to ordinary shape constancy because their stimulus shape was nearly parallel to the picture plane.

# REFERENCES

Ames, A. The illusion of depth from single pictures. J. Opt. Soc. Amer., 1925, 10, 137-148.

Bartley, S. H. A study of the flattening effect produced by optical magnification. Amer. J. Optom., 1951, 28, 290-299.

Bartley, S. H., and Adair, H. J. Comparisons of phenomenal distance in photographs of various sizes. J. Psychol., 1959, 47, 289-295.

Bengston, J. K., Stergios, J. C., Ward, J. L., and Jester, R. E. Optic array determinants of apparent distance and size in pictures. J. Exp. Psychol.: Human Percept. Perform., 1980, 6, 751-759.

Carlbom, I., and Paciorek, J. Planar geometric projections and viewing transformations. Computing Surv., 1978, 10, 465-502.

Cutting, J. E. Perception with an Eye for Motion. Cambridge, MA: Bradford Books/The MIT Press, 1986a.

Cutting, J. E. The shape and psychophysics of cinematic space. Behav. Res. Meth., Instrum., Computers, 1986b, 18, 551-558.

Cutting, J. E. Affine distortions of pictorial space: Some predictions for Goldstein (1987) that La Gournerie (1859) might have made. J. Exp. Psychol.: Human Percept. Perform., 1987, in press.

Ellis, S. R., Tyler, M., Kim, W. S., McGreevy, M. W., and Stark, L. Visual enhancements for perspective displays: Perspective parameters. Proc. 1985 IEEE Intern. Conf. Systems, Man, and Cybernetics, Tucson, AZ, November 12-15, 1985, 815-818.

Enright, J. T. Perspective vergence: oculomotor responses to line drawings. Vision Res., 1987, 27, 1513-1526.

Farber, J. M. The effects of angular magnification on the perception of rigid motion (Doctoral dissertation, Cornell Univ., 1972).

Farber, J., and Rosinski, R. R. Geometric transformations of pictured space. Perception, 1978, 7, 269-282.

Gibson, J. J. Motion Picture Testing and Research (Army Air Forces Aviation Psychology Program Research Reports, No. 7). Washington, DC: GPO, 1947.

Gibson, J. J. The Perception of the Visual World. Boston: Houghton Mifflin, 1950.

Gibson, J. J. A theory of pictorial perception. Audio-Visual Commun. Rev., 1954, 2, 3-23.

Gibson, J. J. Pictures, perspective, and perception. Daedalus, 1960, 89, 216-227.

Gibson, J. J. Ecological optics. Vision Res., 1961, 1, 253-262.

Gibson, J. J. The information available in pictures. Leonardo, 1971, 4, 27-35.

Gibson, J. J. The Ecological Approach to Visual Perception. Boston: Houghton Mifflin, 1979.

Goldstein, E. B. Rotation of objects in pictures viewed at an angle: Evidence for different properties of two types of pictorial space. J. Exp. Psychol.: Human Percept. Perform., 1979, 5, 78-87.

Goldstein, E. B. Spatial layout, orientation relative to the observer, and perceived projection in pictures viewed at an angle. J. Exp. Psychol.: Human Percept. Perform., 1987, 13, 256-266.

Gombrich, E. H. The "what" and the "how": Perspective representation and the phenomenal world. In R. Rudner and I. Scheffler (Eds.), Logic and Art: Essays in Honor of Nelson Goodman. Indianapolis: Bobbs-Merrill, 1972.

Green, R. Determining the preferred viewpoint in linear perspective. Leonardo, 1983, 16, 97-102.

Grunwald, A. J., and Ellis, S. R. Spatial orientation by familiarity cues. Proc. 6th European Conf. Human Decision Making and Cont. Amsterdam: Elsevier Science Publications (in press).

Haber, N. H. Perceiving the layout of space in pictures: A perspective theory based upon Leonardo da Vinci. In C. F. Nodine and D. F. Fisher (Eds.) Perception and Pictorial Representation, Praeger, 1979, 84-89.

Haber, N. H. How we perceive depth from flat pictures. Amer. Scientist, 1980a, 68, 370-380.

Haber, N. H. Perceiving space from pictures: A theoretical analysis. In M. A. Hagen (Ed.), The Perception of Pictures, vol. 1. New York: Academic Press, 1980b, 3-31.

Hagen, M. A. Picture perception: Toward a theoretical model. Psychol. Bull., 1974, 81, 471-497.

Hagen, M. A. Problems with picture perception: A reply to Rosinski. Psychol. Bull., 1976a, 83, 1176-1178.

Hagen, M. A. Influence of picture surface and station point on the ability to compensate for oblique view in pictorial perception. Developmental Psychol., 1976b, 12, 57-63.

Hagen, M. A. Varieties of Realism: Geometries of Representational Art, Cambridge: Cambridge University Press, 1986.

Hagen, M. A., and Elliott, H. B. An investigation of the relationship between viewing condition and preference for true and modified linear perspective with adults. J. Exp. Psychol.: Human Percept. Perform., 1976, 2, 479-490.

Hagen, M. A., and Jones, R. K. Differential patterns of preference for modified linear perspective in children and adults. J. Exp. Child Psychol., 1978, 26, 205-215.

Hay, J. C. The ghost image: A tool for the analysis of the visual stimulus. In R. B. MacLeod and H. L. Pick, Jr. (Eds.), Perception: Essays in Honor of James J. Gibson. Ithaca, NY: Cornell Univ. Press, 1974, 268-275.

Hochberg, J. E. Psychophysics of pictorial perception. Audio-Visual Commun. Rev., 1962, 10, 22-54.

Hochberg, J. E. Some of the things that paintings are. In C. F. Nodine and D. F. Fisher (Eds.) Perception and Pictorial Representation. Praeger, 1979, 17-41.

Jones, R. K., and Hagen, M. A. The perceptual constraints on choosing a pictorial station point. Leonardo, 1978, 10, 1-6.

Kubovy, M. The Psychology of Linear Perspective in Renaissance Art. New York: Cambridge Univ. Press, 1986.

La Gournerie, J. de. Traite de perspective lineaire contenant les traces pour les tableaux plans et courbes, les bas-reliefs, et les decorations theatrales, avec une theorie des effets de perspective. 1 vol. and 1 atlas of plates. Paris: Dalmont et Dunod; Mallet-Bachelier, 1859.

Lumsden, E. A. Perception of radial distance as a function of magnification and truncation of depicted spatial layout. Percept. Psychophys., 1983, 33, 177-182.

Lumsden, E. A. Problems of magnification and minification: An explanation of the distortion of distance, slant, shape, and velocity. In M. A. Hagen (Ed.), The Perception of Pictures, vol. 1, New York: Academic Press, 1980, 91-135.

Mackavey, W. R. Exceptional cases of pictorial perspective. In M. A. Hagen (Ed.), The Perception of Pictures, vol. 1, New York: Academic Press, 1980, 213-223.

McGreevy, M. W., and Ellis, S. R. Direction judgement errors in perspective displays. Proc. 20th Annual Conf. on Manual Contr., 1984, NASA CP 2341, 1, 531-549.

McGreevy, M. W., and Ellis, S. R. The effect of perspective geometry on judged direction in spatial information instruments. Human Factors, 1986, 28, 439-456.

McGreevy, M. W., Ratzlaff, C. R., and Ellis, S. R. Virtual space and two-dimensional effects in perspective displays (manuscript, 1987).

Perkins, D. N. Visual discrimination between rectangular and nonrectangular parallelopipeds. Percept. Psychophys., 1972, 12, 396-400.

Perkins, D. N. Compensating for distortion in viewing pictures obliquely. Percept. Psychophys., 1973, 14, 13-18.

Perkins, D. N. How good a bet is a good form? Perception, 1976, 5, 393-406.

Perkins, D. N. Pictures and the real thing. In P. A. Kolers, M. E. Wrolstad, and H. Bouma (Eds.) Processing of Visible Language 2. New York: Plenum, 1980, 259-278.

Pirenne, M. H. Optics, Painting, and Photography. Cambridge: Cambridge Univ. Press, 1970.

Pirenne, M. H. Vision and art. In E. C. Carterette and M. P. Friedman (Eds.) Handbook of Perception, vol. 5, Seeing. New York: Academic Press, 1975, 433-490.

Purdy, W. C. The hypothesis of psychophysical correspondence in space perception. Ithaca, NY: General Electric Advanced Electronics Center, September 1960 (NTIS R60ELC56).

Rogers, S. Representation and reality: Gibson's concept of information and the problem of pictures. (Doctoral dissertation, Royal College of Art, 1985).

Rosinski, R. R. Picture perception and monocular vision: A reply to Hagen. Psychol. Bull., 1976, 83, 1172-1175.

Rosinski, R. R., and Farber, J. Compensation for viewing point in the perception of pictured space. In M. A. Hagen (Ed.), The Perception of Pictures, vol. 1, New York: Academic Press, 1980, 137-176.

Rosinski, R. R., Mulholland, T., Degelman, D., and Farber, J. Picture perception: An analysis of visual compensation. Percept. Psychophys., 1980, 28, 521-526.

Schlosberg, H. Stereoscopic depth from single pictures. Amer. J. Psychol., 1941, 54, 601-605.

Sedgwick, H. A. The geometry of spatial layout in pictorial representation. In M. A. Hagen (Ed.), The Perception of Pictures, vol. 1, New York: Academic Press, 1980, 33-90.

Sedgwick, H. A. Environment-centered representation of spatial layout: Available visual information from texture and perspective. In J. Beck, B. Hope, and A. Rosenfeld (Eds.), Human and Machine Vision. New York: Academic Press, 1983.

Sedgwick, H. A. Space perception. In K. Boff, L. Kaufman, and J. Thomas (Eds.), Handbook of Perception and Human Performance (Vol. 1). New York: Wiley, 1986.

Sedgwick, H. A. A production system modeling high-level visual perspective information for spatial layout. Technical Report No. 298, New York Univ. Depart. of Computer Sci., 1987a.

Sedgwick, H. A. Layout2: A production system modeling visual perspective information. Proc. IEEE First Intern. Conf. on Computer Vision, London, June 8-11, 1987b.

Sedgwick, H. A., and Levy, S. Environment-centered and viewer-centered perception of surface orientation. Computer Vision, Graphics, and Image Processing, 1985, 31, 248-260.

Smith, O. W. Comparison of apparent depth in a photograph viewed from two distances. Percept. Motor Skills, 1958a, 8, 79-81.

Smith, O. W.  Judgments of size and distance in photographs.  <u>Amer. J. Psychol.</u>, 1958b, <u>71</u>, 529-538.

Smith, O. W., and Gruber, H.  Perception of depth in photographs.  <u>Percept. Motor Skills</u>, 1958, <u>8</u>, 307-313.

Smith, O. W., Smith, P. C., and Hubbard, H.  Perceived distance as a function of the method of representing perspective.  <u>Amer. J. Psychol.</u>, 1958, <u>71</u>, 662-674.

Smith, P. C., and Smith, O. W.  Ball throwing responses to photographically portrayed targets.  <u>J. Exp. Psychol.</u>, 1961, <u>62</u>, 223-233.

Wallach, H.  The apparent rotation of pictorial scenes.  In M. Henle (Ed.) <u>Vision and Artifact</u>, New York:  Springer, 1976, 65-69.

Wallach, H.  Perceiving a stable environment.  <u>Scientific Amer.</u>, 1985, <u>252</u>, 118-124.

Wallach, H., and Marshall, F. J.  Shape constancy in pictorial representation.  <u>Percept. Psychophys.</u>, 1986, <u>39</u>, 233-235.

Witkin, A. P.  Shape from contour.  MIT technical report AI-TR-589, November 1980.
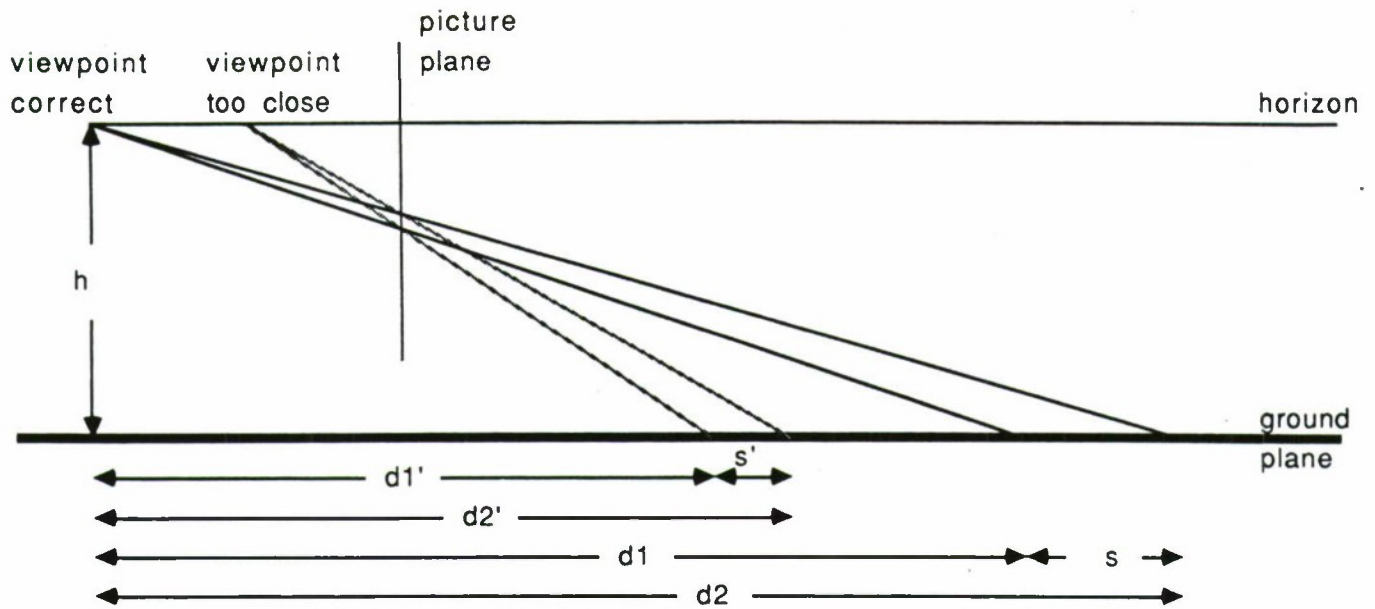
Figure 1.- Close viewing compresses geometrically specified virtual depth.
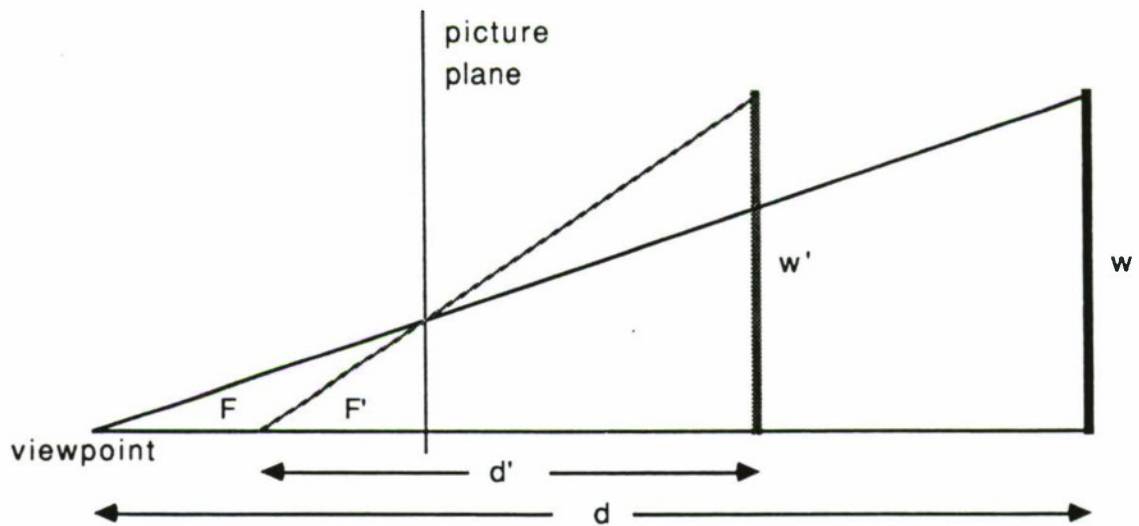


Figure 2.- Close viewing leaves geometrically specified virtual frontal dimensions unchanged.

pictorial space

picture plane

real space

viewpoint correct

viewpoint too close
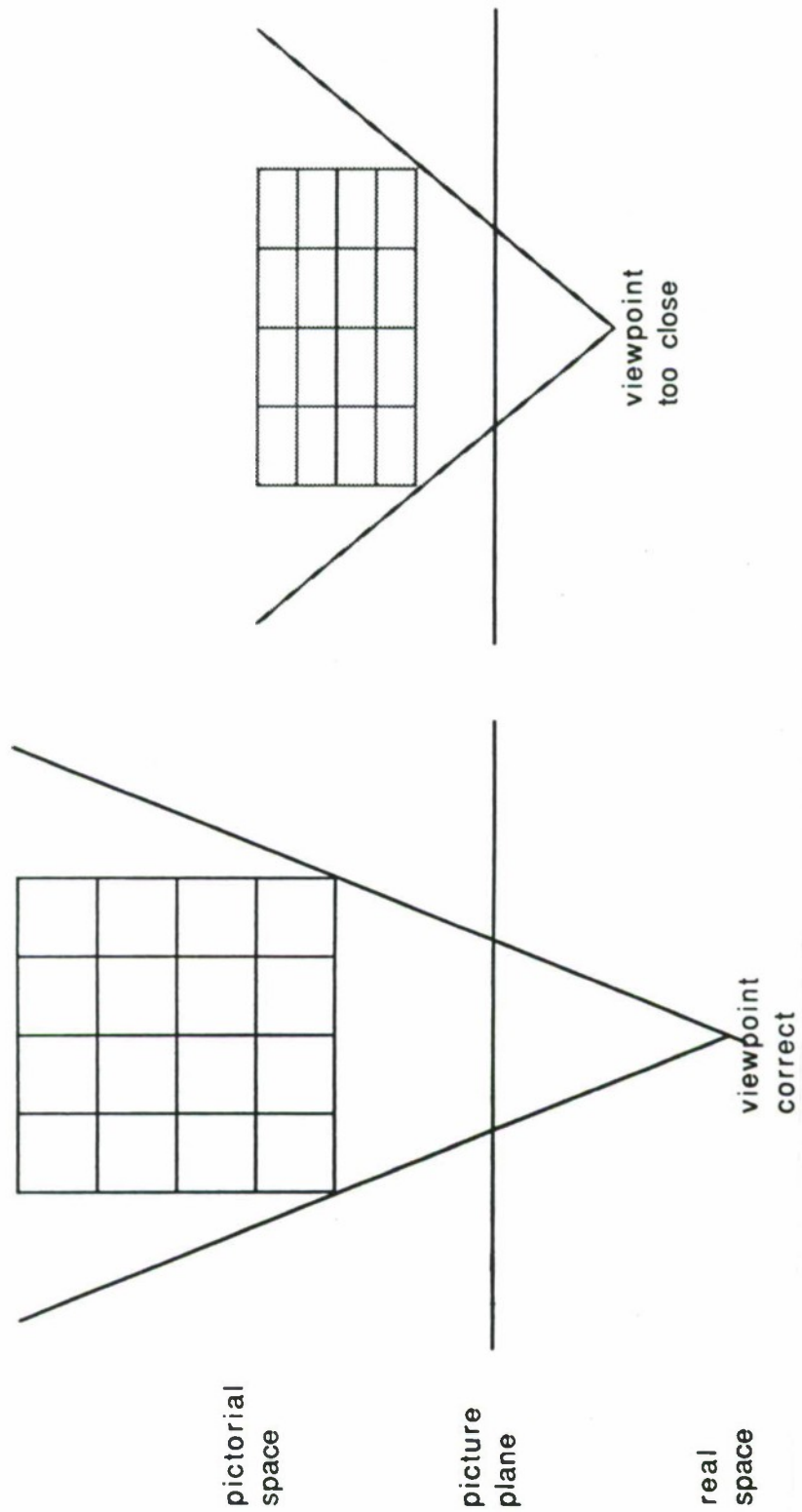
Figure 3.- Close viewing distorts geometrically specified virtual shape.

pictorial space

picture plane

real space

VP

VP

viewpoint correct

VP

VP

viewpoint too close
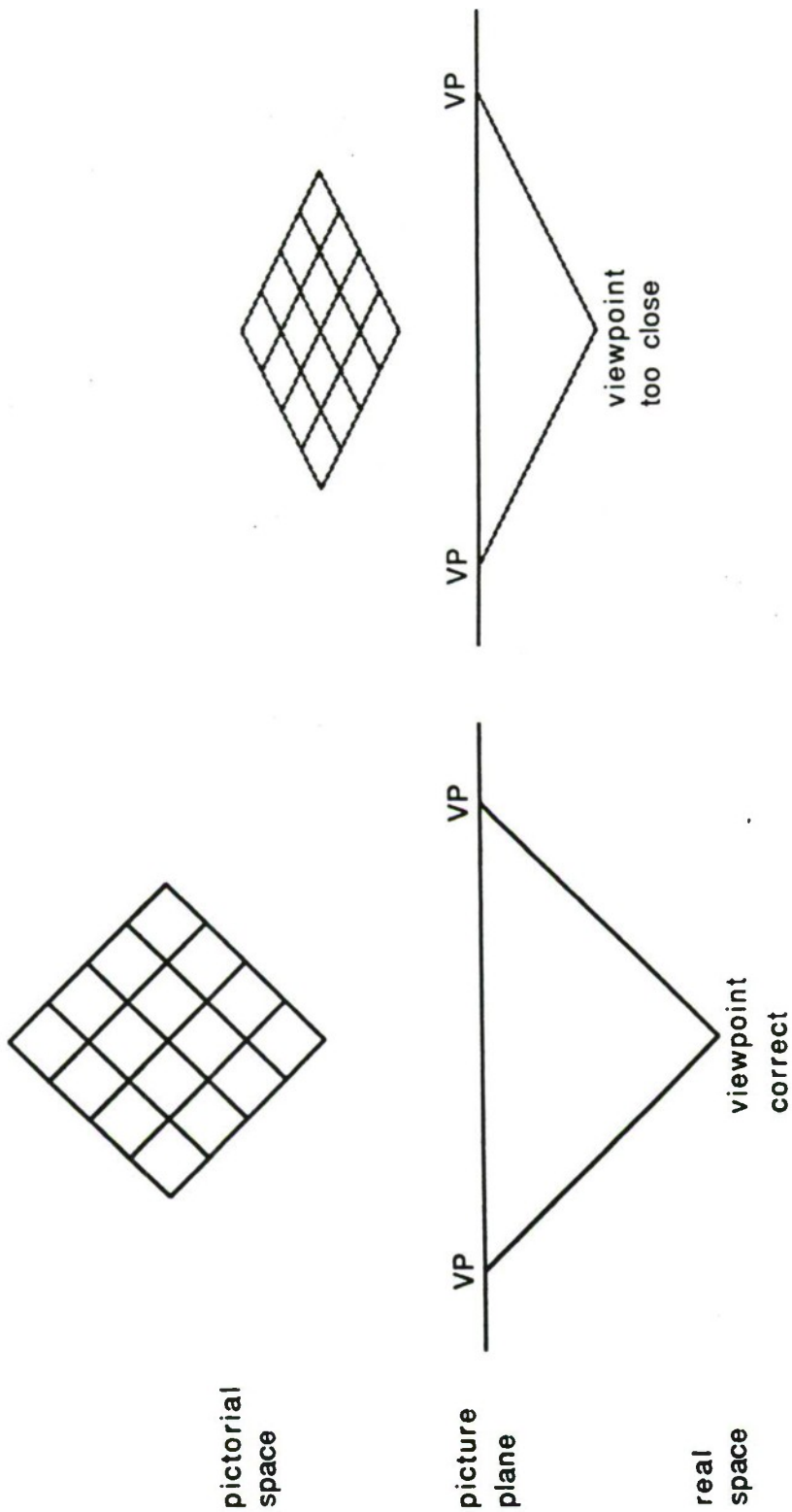
Figure 4.- Vanishing points geometrically specify distortions in virtual orientation with close viewing.

15-20
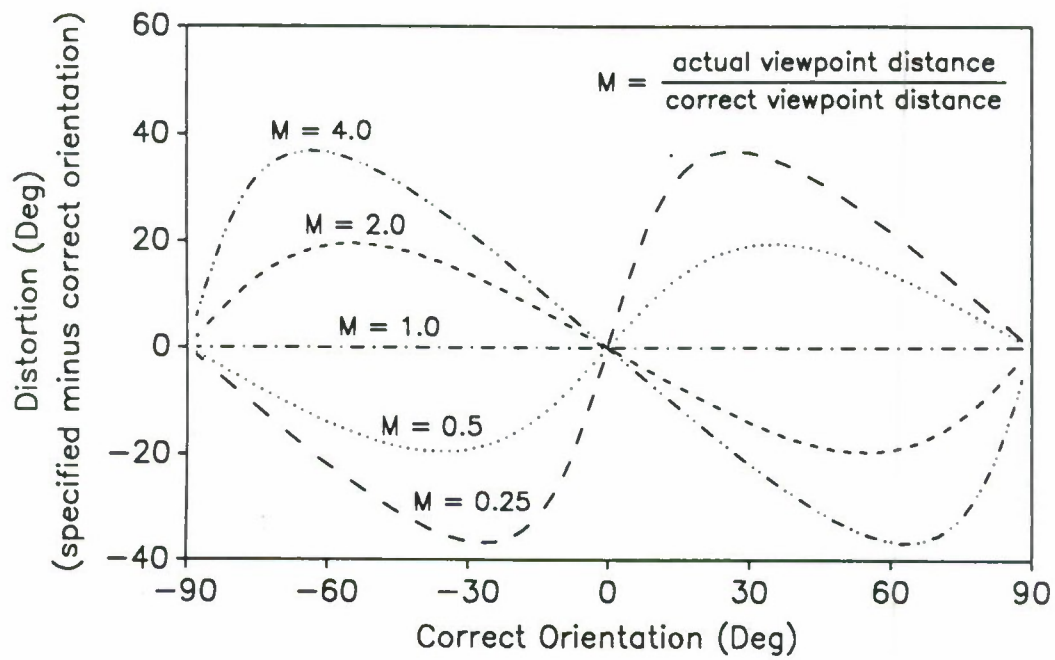
Figure 5.- Geometrically specified distortion in virtual orientation as a function of viewing distance.
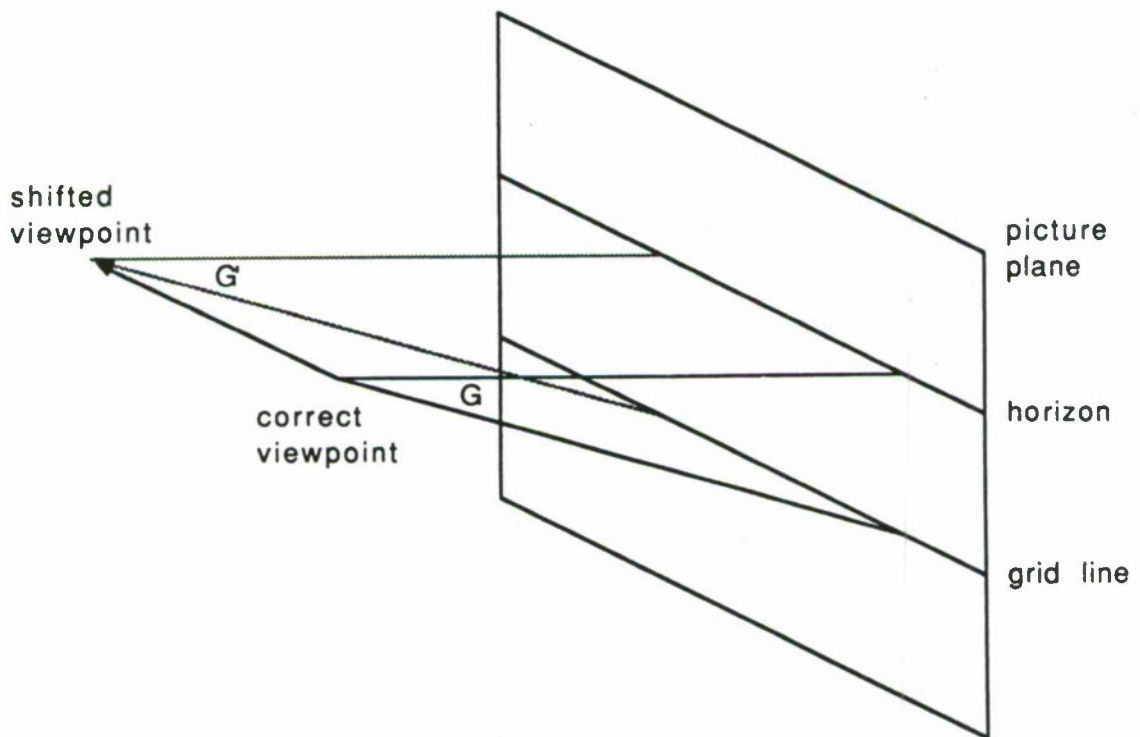


Figure 6.- Lateral shifts in viewpoint do not change geometrically specified virtual depth.
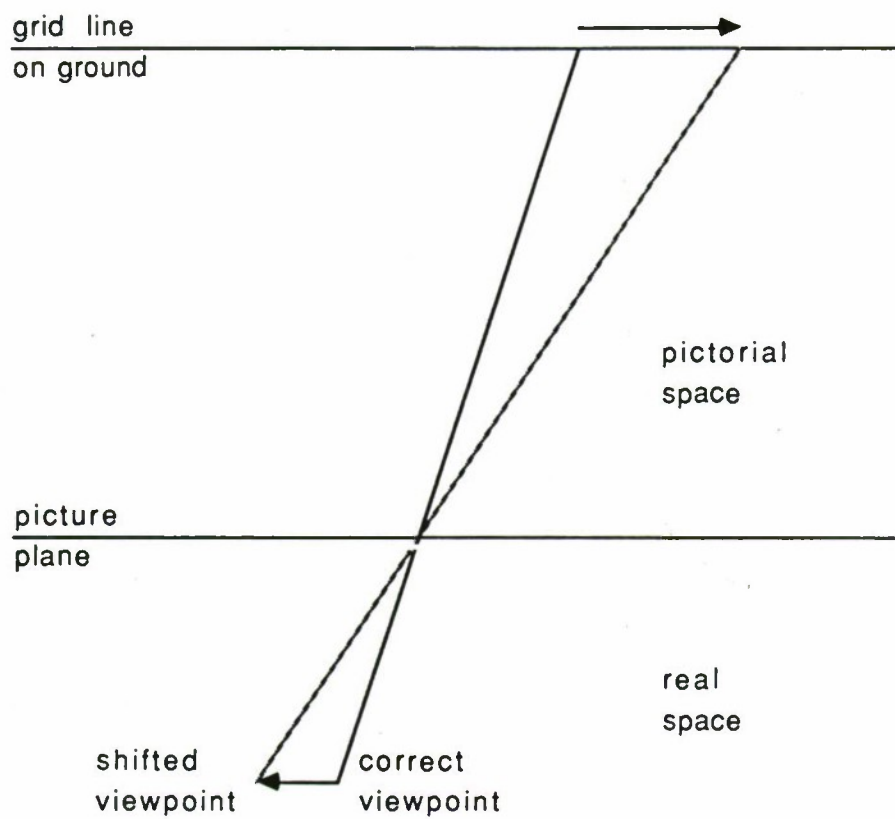
Figure 7.- Lateral shifts in viewpoint geometrically specify lateral shifts in virtual space.

pictorial
space

picture
plane

shifted          correct
viewpoint        viewpoint

real
space
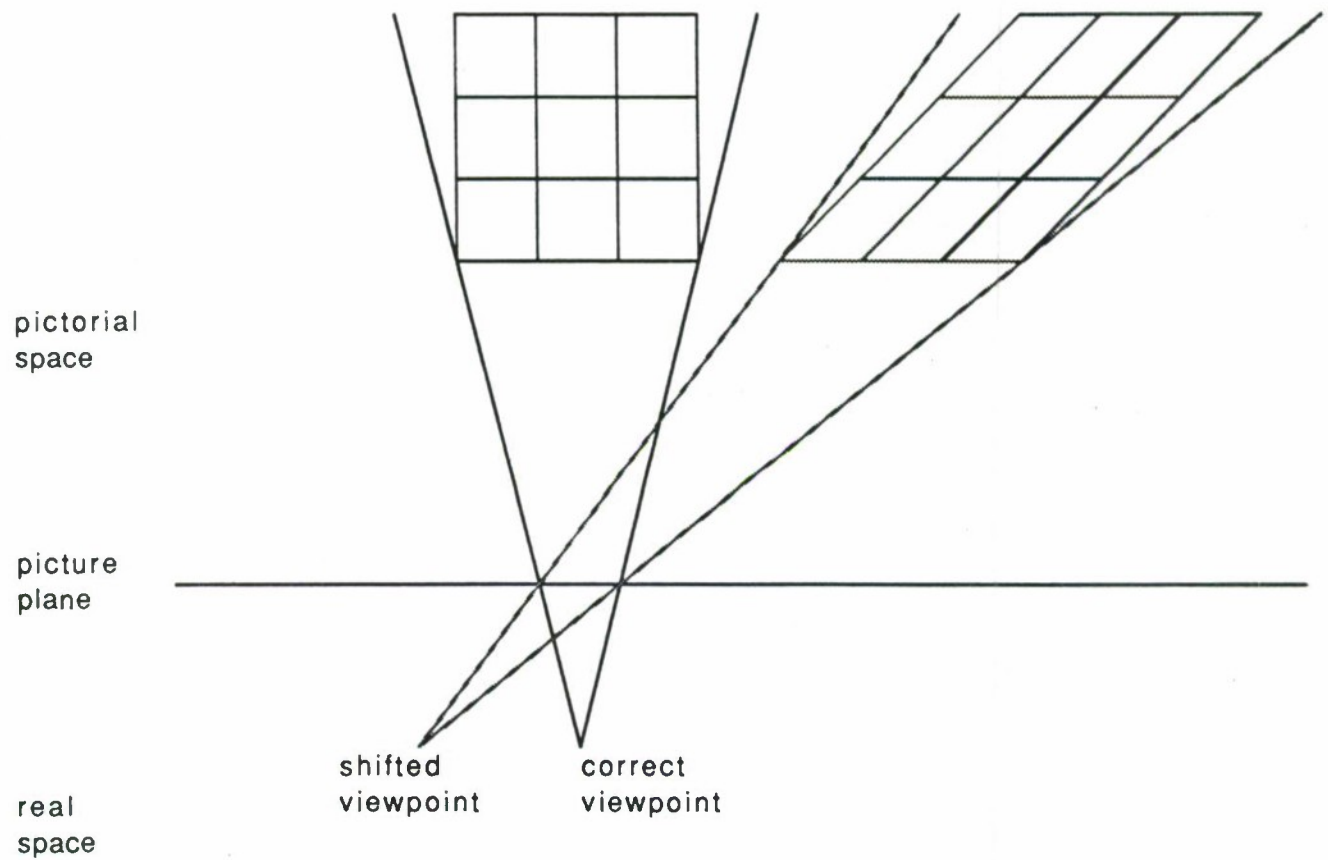
Figure 8.- Lateral shifts in viewpoint geometrically specify a shearing of virtual space.
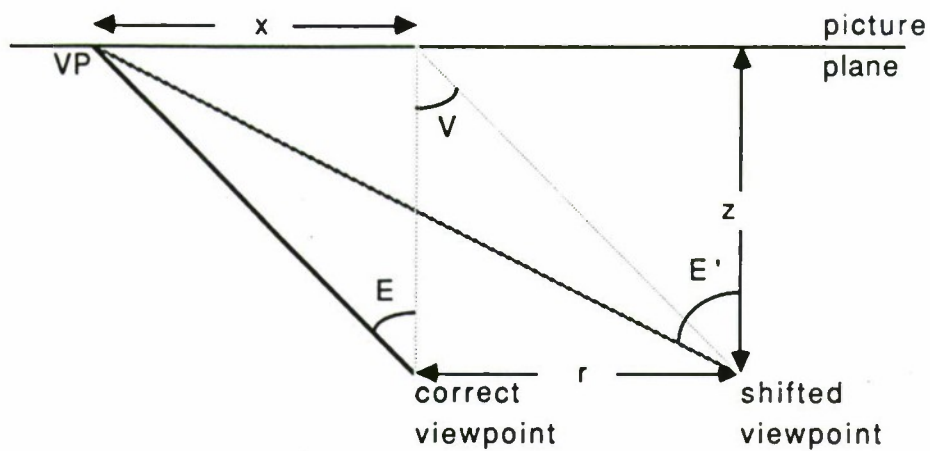
Figure 9.- Vanishing points geometrically specify distortions in virtual orientation with lateral shifts in viewpoint.
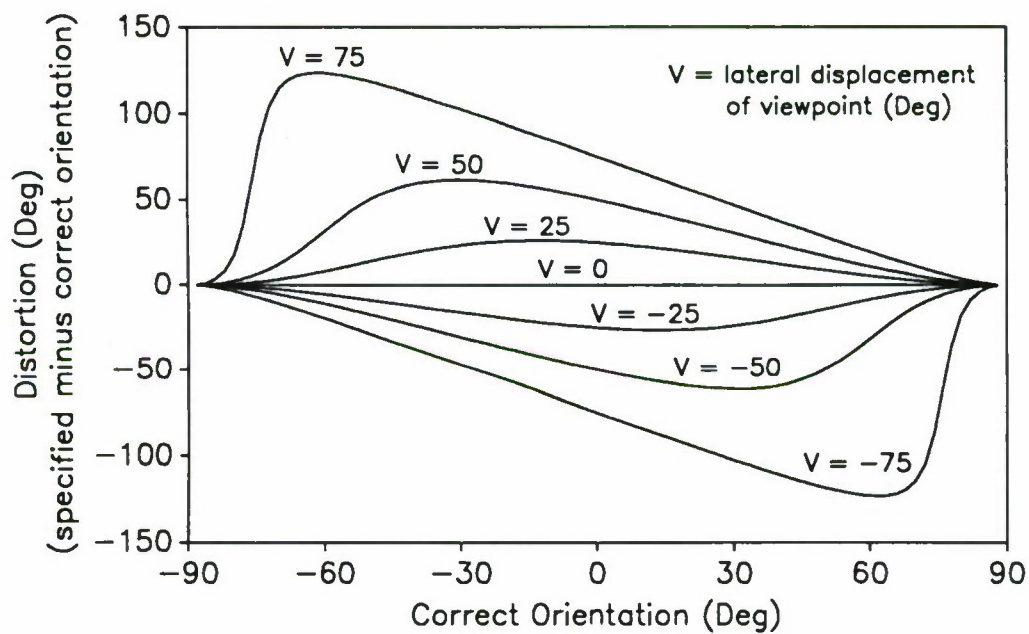


Figure 10.- Geometrically specified distortion in virtual orientation as a function of lateral shift in viewpoint.

# PERCEIVED ORIENTATION, SPATIAL LAYOUT AND THE GEOMETRY OF PICTURES

E. Bruce Goldstein
Department of Psychology
University of Pittsburgh
Pittsburgh, Pennsylvania

The purpose of this paper is to discuss the role of geometry in determining the perception of spatial layout and perceived orientation in pictures viewed at an angle. This discussion derives from Cutting's (1988) suggestion, based on his analysis of some of my data (Goldstein, 1987), that the changes in perceived orientation that occur when pictures are viewed at an angle can be explained in terms of geometrically produced changes in the picture's virtual space. Before dealing with Cutting's idea, let's first consider the paper that stimulated it.

Goldstein (1987) distinguishes between three different perceptual attributes of pictures:

1. Perceived orientation. The direction a pictured object appears to point when extended out of a picture, into the observer's space.

2. Perceived spatial layout. The perception of the layout in three-dimensional space of objects represented in the picture.

3. Perceived projection. The perception of the projection of the picture's image on the observer's retina.

One basis for making these distinctions is that the perception of these attributes is affected differently by changes in the observer's viewing angle. Perceived orientation and perceived spatial layout, the two attributes we will focus on in this paper, differ in the following way:

1. Perceived spatial layout remains relatively constant with changes in viewing angle. This "layout constancy" is demonstrated by presenting photographs of triangular arrays of dowels like the ones in figure 1, and asking subjects to reproduce the layout this array would have if viewed from directly above. The results of these experiments, indicated by the general correspondence between the shapes of the solid triangles in figure 2, indicate that changing viewing angle causes only small changes in a subject's ability to reproduce spatial layout. This relative constancy has also been observed for other arrays and for pictures of environmental scenes (Goldstein, 1979, 1987).

2. Perceived orientation, on the other hand, undergoes large changes with changes in viewing angle. Figure 3 shows the average perceived orientations for four observers judging the orientations defined by pairs of dowels BA and BC of figure 1. When the picture is viewed at an angle of 20° (far to the right side of the picture plane), the relationship between the two orientations is different than when it is viewed at 160° (far to the left side of the picture plane). These differences are manifestations of the *differential rotation effect*—the fact that pictured objects oriented more parallel to the picture plane rotate less in response to an observer's change in viewing angle than do pictured objects that are oriented more perpendicular to the picture plane.

(See Goldstein, 1979, 1987, for a more detailed graphical presentation of similar data for a number of viewing angles).

In my paper I presented evidence that the subject's awareness of the picture plane is one of the causes of these changes in the perceived orientation of different objects relative to one another. Cutting (1988) has offered an alternate explanation–that perceived orientation is controlled by the geometrical changes associated with the affine shear that accompanies changes in viewing angle. His analysis is based on an analysis of the virtual space defined by a picture–that is, the three-dimensional space that corresponds to the picture's geometrical array. Cutting's original analysis was based on a formula developed by Rosinski et al. (1980), but it is also possible to use the graphical method illustrated in the top part of figure 4 (see Cutting, 1986, p. 36, for an illustration of the geometrical method used to construct this figure) to determine how the picture's virtual space is affected by changes in viewing angle. This figure shows the virtual space defined by the array in the center top of the figure, for viewing angles of 20°, 90°, and 160°.

After determining the virtual space defined by my triangular array at different viewing angles, Cutting used the orientations defined by this space to predict perceived orientations at each viewing angle. The resulting predictions for perceived orientations fit the data well at some viewing angles and not as well at others. Consider, for example, his prediction for a viewing angle of 160°. We can compare the predicted orientations shown at the top right of figure 4 to those determined empirically by constructing a triangle based on the empirically determined perceived orientations. Such a triangle, calculated from the data in figure 3 of Goldstein (1987)[1] and shown on the lower right of figure 4, is oriented slightly differently than Cutting's predicted triangle, but has the same general shape. The fit is not, however, as good for a viewing angle of 20°; at that angle Cutting's predicted orientations for the directions defined by B → C and C → A differ from those determined empirically.

Although these differences between geometrically predicted and empirical results suggest that geometry cannot supply the entire explanation for the changes in perceived orientation that occur with changes in viewing angle, Cutting's model does succeed in predicting the differential rotation effect. Geometry may, therefore, play at least some role in determining perceived orientation, and it is this role I wish to focus on now.

Let's assume for the moment that perceived orientations *are* linked to the changes that occur in virtual space with changes in viewing angle. This possible linkage between changes in virtual space and perceived orientation becomes particularly significant when we consider that *these same changes in virtual space* cause little change in the observer's perception of spatial layout. This constancy of spatial layout occurs not only for changes in viewing angle, as illustrated by the solid triangles in figure 2, but also for changes in viewing distance, as indicated by comparing the solid and dashed triangles in figure 2. The solid triangles were produced by subjects viewing the array in figure 1 from a distance of 8 in., whereas the dashed triangles were produced from a viewing distance of 64 in. Despite this eight-fold difference in distance, which causes a large expansion of virtual space,[2] there are only small differences between the triangles.

---

[1]The data on which these triangles are based were collected using a stimulus with the same layout as the stimulus shown in figure 1, but the photograph of the dowels was taken from a slightly lower angle (see Goldstein, 1987, for a picture of this stimulus).

[2]The use of the graphical method to determine how virtual space is changed by this increase in distance indicates that the expansion of the space caused by changing the viewing distance from 8 to 64 in. produces an elongated triangle in which side BA is stretched to four times the length of side BC.

What we have here, therefore, is a situation in which large changes in virtual space cause little or no change in the perception of spatial layout, but which, to the extent that the geometrical hypothesis is correct, cause large changes in perceived orientation. This situation raises the possibility that perceived orientation may result directly from stimulus geometry, whereas the perception of spatial layout may involve a processing step to compensate for the geometrical changes caused by viewing at an angle.

This idea of a compensation mechanism is not new. Pirenne (1970), Rosinski, et al. (1980) and Kubovy (1986) have linked such mechanisms to the subject's awareness of the picture plane; however, the exact operation of this compensation mechanism has never been specified. The first question that should be asked to help elucidate the nature of this hypothetical mechanism is: What stimulus manipulation will cause a subject's perception of layout to correspond to the picture's virtual space–or, put another way, What stimulus manipulation will eliminate layout constancy?

It is also possible that layout constancy is the outcome, not of a compensation mechanism, but of the subject's attention to information in the picture that remains invariant with changes in virtual space. While it is easy to talk glibly about invariant information, we need to identify this information if, in fact, it exists.

Finally, returning to perceived orientation, the suggestion that this percept may result directly from stimulus geometry cannot be the whole story. It seems clear that the observer's awareness of the angle of view is also important (Goldstein, 1987), although exactly how this factor interacts with stimulus geometry remains to be determined.

Obviously, many questions remain to be answered before we fully understand the mechanisms underlying perceived orientation and perceived spatial layout. These questions are important, not only because they suggest possibilities for future research that could yield answers that will greatly enhance our understanding of picture perception, but also because they acknowledge an important fact about picture perception: Perceived orientation and perceived spatial layout are affected differently by changes in viewing angle, are probably controlled by different mechanisms, and should, therefore, be clearly distinghished from one another in future research on picture perception.

# REFERENCES

Cutting, J.E. (1986) Perception with an eye for motion.  Cambridge: MIT Press.

Cutting, J.E. (1988) Affine distortions of pictorial space: Some predictions for Goldstein (1987) that LaGournerie (1859) may have made. J. Exp. Psychol.: Human Percept. Perform., in press.

Goldstein, E.B. (1979) Rotation of objects in pictures viewed at an angle: Evidence for different properties of two types of pictorial space.  J. Exp. Psychol.: Human Percept. Perform. 5, 78-87.

Goldstein, E.B. (1987) Spatial layout, orientation relative to the observer, and perceived projection in pictures viewed at an angle.  J. Exp. Psychol.: Human Percept. Perform. 13, 256-266.

Kubovy, M. (1986) The psychology of perspective and Renaissance art. Cambridge: Cambridge Univ. Press.

Pirenne, M.H. (1970) Optics, painting, and photography. Cambridge, Cambridge Univ. Press.

Rosinski, R.R., Mulholland, T., Degelman, D., & Farber, J. (1980) Picture perception: An analysis of visual compensation. Percept. Psychophys. 28, 521-526.
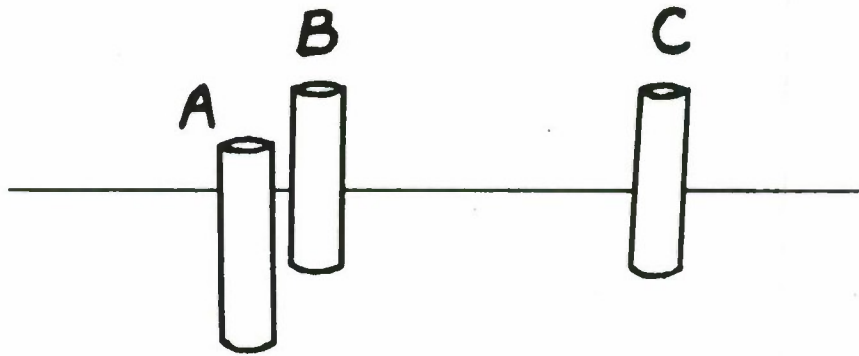
Figure 1.– Stimulus used to determine the perceived spatial layouts of figure 2 and the perceived orientations in figure 3. In the actual photographic stimuli the dowels had horizontal black and white stripes to clearly distinguish them from the background.



160°         90°         20°
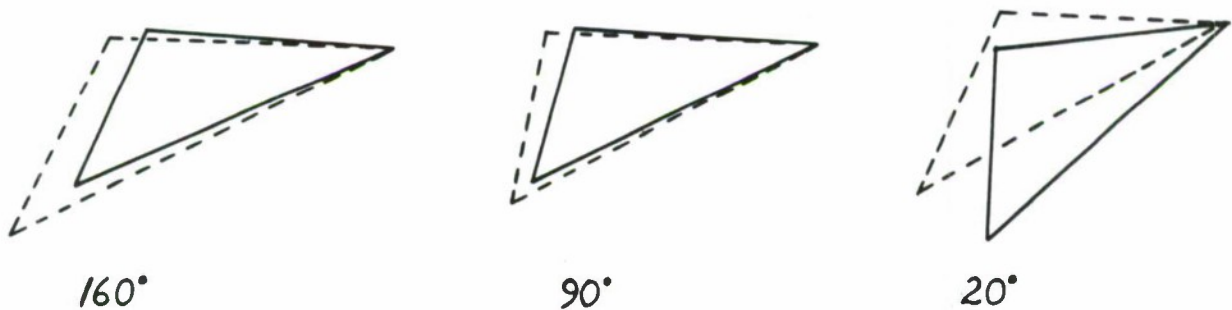
Figure 2.– Solid triangles—average spatial layouts produced by four observers viewing the array of rods in figure 1 from a distance of 8 in. at viewing angles of 20°, 90°, and 180°. Dashed triangles–average spatial layouts produced by the same observers from a viewing distance of 64 in. Viewing angle is the angle between the observer's line of sight and the picture plane, with a viewing angle of 0° occurring when the observer is looking at the right edge of the picture and a viewing angle of 180°, occurring when the observer is looking at the left edge. (See Goldstein (1987) for further details of stimulus specification and procedures.)
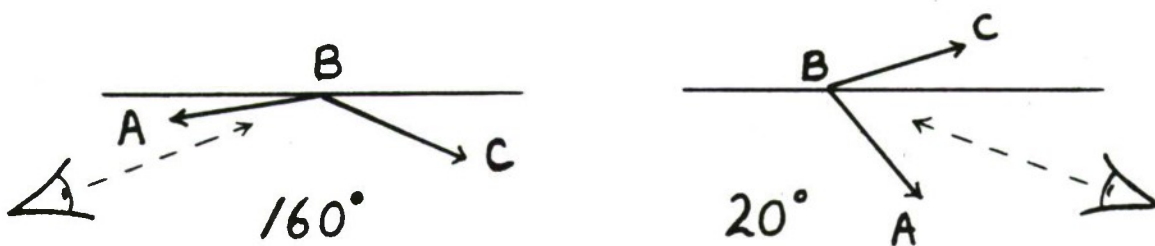
Figure 3.– Averaged perceived orientations defined by dowels BA and BC of figure 1, when viewed at viewing angles of 20° and 160°. The picture plane is indicated by the horizontal line and the observer's position is shown by the schematic eye. Perceived orientations are indicated by the direction of the arrows. Note that for a viewing angle of 20°, the orientation of BC points behind the picture plane. This is a typical result, which has been previously reported (Goldstein, 1979, 1987).

AFFINE TRANSFORMS

CALCULATED FROM
PERCEIVED ORIENTATIONS
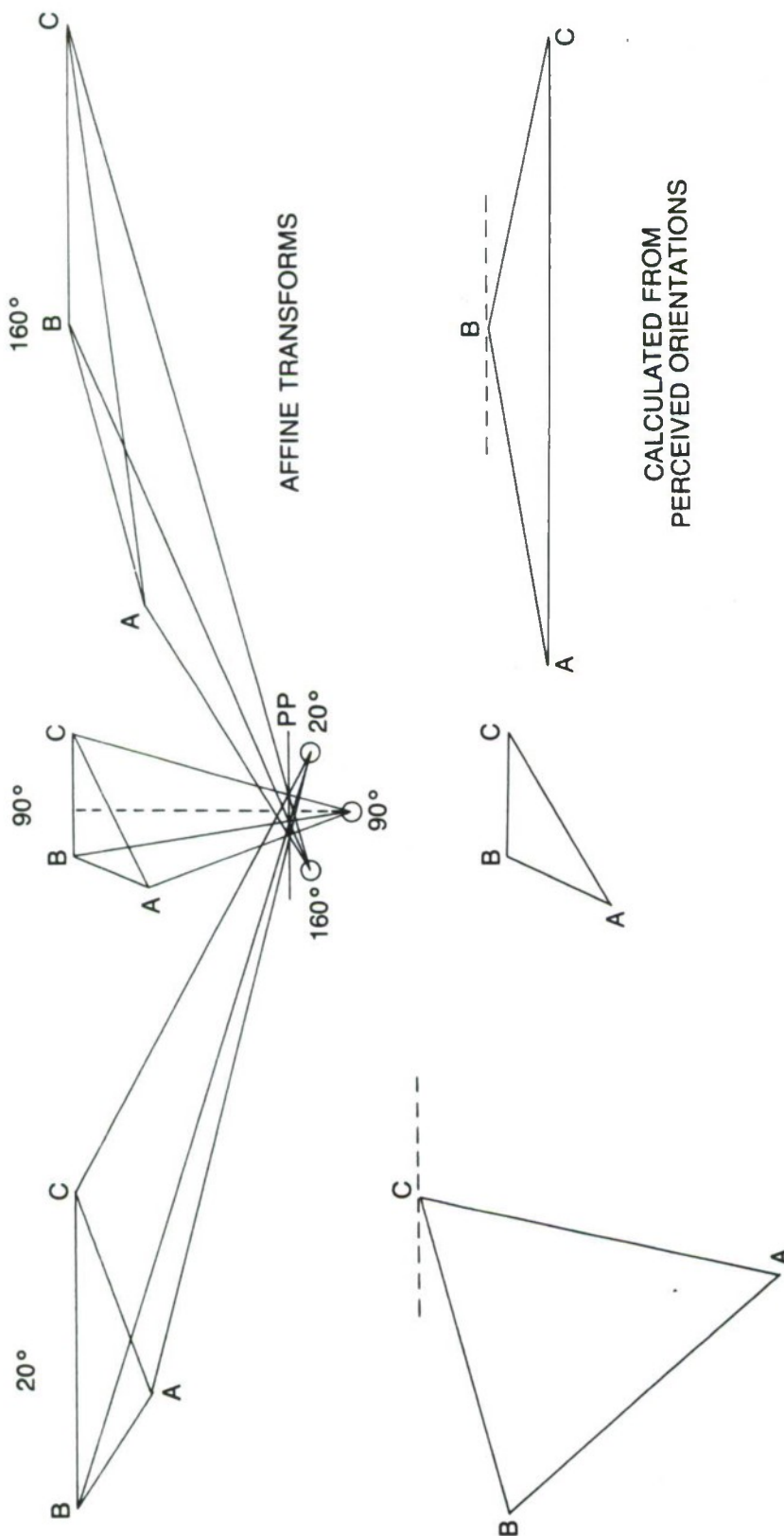
Figure 4.– Top: center–layout of the triangular array used in Goldstein (1987); left–affine transformed array for a viewing angle of 20°; right–affine transformed array for a viewing angle of 160°. PP = picture plane; open circles = positions of observers at viewing angles of 20°, 90°, and 160° relative to the PP. The dashed vertical line is the observer's line of sight for the 90° viewing angle. Bottom: layouts calculated from the empirically determined perceived orientations in figure 3 of Goldstein (1987). Since perceived orientations do not provide information regarding size, the sizes of these triangles were determined by setting the length of side BC equal to the length of side BC of the corresponding triangle above. The triangles were constructed by drawing each line so its orientation matches its empirically determined perceived orientation relative to the picture plane. The orientation of the picture planes for the lower triangles are indicated by dashed lines for the 20° and 160° viewing angles. The picture plane is omitted for the 90° viewing angle for clarity, since the angle between BC and the picture plane is 2°.

16-7

# ON THE EFFICACY OF CINEMA, OR
# WHAT THE VISUAL SYSTEM DID NOT EVOLVE TO DO

James E. Cutting
Department of Psychology
Cornell University, Ithaca, New York

My topic concerns spatial displays, and a constraint that they do not place on the use of spatial instruments. Much of the work done in visual perception by psychologists and by computer scientists has concerned displays that show the motion of rigid objects. Typically, if one assumes that objects are rigid, one can then proceed to understand how the constant shape of the object can be perceived (or computed) as it moves through space. Many have assumed that a rigidity principle reigns in perception; that is, the visual system prefers to see things as rigid. There are now ample reasons to believe, however, that a rigidity principle is not always followed. Hochberg (1986), for example, has outlined some of the conditions under which a rigid object ought to be seen, but is not. Some of these concern elaborations of some of the demonstrations that Adelbert Ames provided us more than 35 years ago.

There is another condition of interest with respect to rigidity and motion perception. That is, not only must we know about those situations in which rigidity ought to be perceived, but is not, we also must know about those conditions in which rigidity ought not to be perceived, but is. Here I address one of these conditions, with respect to cinema. But before discussing cinema, I must first consider photography.

When we look at photographs or representational paintings, our eye position is not usually fixed. A puzzle arises from this fact: Linear perspective is mathematically correct for only one station point, or point of regard, yet almost any position generally in front of a picture will do for object identity and layout within the picture to appear relatively undisturbed. Preservation of phenomenal identity and shape of objects in slanted pictures is fortunate. Without them the utility of pictures would be vanishingly small. Yet the efficacy of slanted pictures is unpredicted by linear perspective theory.

This puzzle was first treated systematically by La Gournerie in 1859 (see Pirenne, 1970). I call it La Gournerie's paradox; Kubovy (1986) has called it the robustness of perspective. The paradox occurs in two forms: The first concerns viewing pictures either nearer or farther than the proper station point; the second and more dramatic concerns viewing pictures from the side. Both are shown in the top panels of figure 1.

To consider either distortion one must reconstruct, as La Gournerie did, the geometry of pictured (or virtual space) behind the picture plane. The premise for doing so is that the image plane is unmoving, but invisible, and that observers look through it into pictured space to make sense out of what is depicted. Invisibility is, in many cases, obviously a very strong, if not false, assumption, but it yields interesting results. Possible changes in viewing position are along the z axis, orthogonal to the picture plane, and along the x or y axes, parallel to it. Both generate affine transformations in depth in all xz planes of virtual space. Observer movement along x or y axes

also generates perspective transformations of the image, but these will not be considered here (Cutting 1986a, 1986b).

In the upper left panel, four points are projected onto the image plane as might be seen in a large photograph taken with a short lens. When the observer moves closer to the image, as in the upper middle panel, the projected points must stay in the same physical locations in the photo. Thus, the geometry of what lies behind must change. Notice that the distance between front and back pairs of points of this four-point object is compressed, a collapse of depth like that when looking through a telephoto lens. All changes in z axis location of the observer create compression or expansion of the object in virtual space. When an observer moves to the side, as seen in the upper right panel, points in virtual space must shift over, and do so by different amounts. Such shifts are due to affine shear. All viewpoints of a picture yield additive combinations of these two affine effects--compression (or expansion) and shear.

Such effects are compounded when viewing a motion sequence, as shown in the lower panels of figure 1. In particular, an otherwise rigid object should appear to hinge and become nonrigid over the course of several frames for a viewer seated to the side. Theoretically, the problem this poses for the cinematic viewer is enormous--every viewer in a cineauditorium has an eye position different than the projector and camera position, and thus, by the rules of perspective, no moving object should ever appear rigid. This is, I claim, the fundamental problem of the perception of film and television.

Most explanations for the perception of pictures at a slant are in sympathy with Helmholtz. Pirenne (1970, p. 99), for example, suggested that "an unconscious intuitive process of psychological compensation takes place, which restores the correct view when the picture is looked at from the wrong position." Pirenne's unconscious inference appears to unpack the deformations through some process akin to mental rotation (Shepard & Cooper, 1982). According to this view, the mind detransforms the distortions in pictured space so that things may be seen properly, and although Pirenne didn't discuss film, it might hold equally for film seen from the front row, side aisle. The force of my presentation is to show that this view is not necessary in the perception of slanted cinema. But first consider how this account might proceed.

Pirenne and others have suggested at least three sources of image surface information that might be used to "correct" slanted images--(1) the edges of the screen, which yield a trapezoidal frame of reference; (2) binocular disparities, which grade across the slanted surface; and (3) projection surface information such as texture and specularities. Since I am interested in none of these, I removed them from my displays through a double projection scheme, as shown in figure 2. If one considers the situation of viewing slanted cinema, one has the real, slanted surface and one can measure a cross section of that optic array from it. This would be an imaginary projection surface. Once considered this way, one can reverse the two, placing the real surface in front of the imaginary, and this is what I did.

In this manner, although the display frame was always rectangular for the observer, the shapes of rotating stimuli were like those seen from the side, with the right-edge elements in each frame longer than the left-edge ones and with the z axis compressed. This simulation yields a perspective transformation of the image screen, and a nonperspective transformation of the stimulus behind it in virtual space. I presented viewers with computer-generated, rotating, rectangular solids. Two factors are relevant to this discussion. (For a more complete analysis see Cutting, 1987.)

First, half the solids presented were rigid, half nonrigid. Nonrigid solids underwent two kinds of transformation during rotation—one affine, compressing and expanding the solid like an accordion along one of its axes orthogonal to the axis of rotation during rotation, and one non-affine, with a corner of the solid moving through the same excursion. Deformations were sinusoidal and were accomplished within one rotation of the stimulus. It was relatively easy to see the large excursions as making the solid nonrigid; it was more difficult in smaller excursions. This nonaffine deformation was much easier to see than the affine deformation, but there were no interactions involving types of nonrigidity, so here I will collapse across them (see Cutting, 1987, for their separate discussion).

Second, stimuli were presented with cinematic viewpoint varied; in Experiment 1, half were projected as if viewed from the correct station point, half as if seen from the side, with the angle between imaginary and real projections surfaces set at 23°. The latter condition allows investigation of La Gournerie's paradox, and compounds the nonrigid deformations of the stimulus in pictorial space with an additional perspective transformation of the image.

Viewers looked at many different tokens of all stimuli, and used a bipolar graded scale of rigidity and confidence, from 1 to 9—with 1 indicating high confidence in nonrigidity, 9 high confidence in rigidity, and 5 indicating no confidence either way.

Figure 3 shows the results of the first experiment for rigid and nonrigid stimuli, at both 90° and simulated 67° viewing angles. Two effects are clear. First, rigid stimuli were seen as equally rigid regardless of simulated viewpoint in front of the screen, and second, nonrigid stimuli were seen as equally nonrigid regardless of simulated viewpoint.

The lack of difference in the slanted and unslanted simulated viewing conditions is striking, but it could be due to the fact that the screen slant was relatively slight. Experiment 2, then, introduced a third viewing condition, a steeper angle—45°. A fourth condition was also introduced. Its impetus came from structure-from-motion algorithms in machine vision research. Several people suggested to me that screen slant could be another parameter in rigidity-finding algorithms and that only a few more frames or points might be needed to specify slant. To test for this idea, I introduced a variable screen-slant condition, where the simulated slant of the screen oscillated between 80° and 55°, with a mean of 67°. It seemed highly unlikely that an algorithm could easily solve for both rigidity and a dynamically changing projection surface.

This time stimuli were generated in near-parallel and polar perspective. Again, stimuli could be rigid or nonrigid. Selected results for the nonrigid stimuli are shown in figure 4, and show two striking effects. First, the variable 67° screen slant condition was not different from the nonvarying condition, and the lack of difference would seem to be embarrassing for any structure-through-motion approach to the perception of these stimuli that includes screen slant as a variable to be solved for. Second, if simulated screen slant is great enough, all stimuli begin to look nonrigid.

A more interesting result is an interaction concerning near-parallel and polar projected stimuli, as shown in figure 5, with the two 67° conditions collapsed, and all rigid and nonrigid trials collapsed. The near-parallel projected stimuli show no difference in perceived rigidity from any angle that they are viewed; the more polar projected stimuli, on the other hand, show a sharp decrease in perceived nonrigidity as the angle or regard increases.

This latter effect adds substance to other results in the literature. For example, Hagen and Elliott (1976) found what they called a "zoom effect"—the general preference for static stimuli seen is more parallel than polar projection. Here, in cinematic displays, stimuli that are near-parallel-projected are seen as more rigid from more places in a cineauditorium.

In conclusion, let us be reminded that photographs and cinema are visual displays that are also powerful forms of art. Their efficacy, in part, stems from the fact that, although viewpoint is constrained when composing them, it is not nearly so constrained when viewing them. The reason that viewpoint is relatively unconstrained, I claim, is not that viewers "take into account" the slant of the screen, but that the visual system does not seem to compute the relatively small distortions in the projections, at least for certain stimuli that are projected in a near-parallel fashion.

It is obvious that our visual system did not evolve to watch movies or look at photographs. Thus, what photographs and movies present to us must be allowed in the rule-governed system under which vision evolved. Slanted photographs and cinema present an interesting case where the rules are systematically broken, but broken in a way that is largely inconsequential to vision. Machine-vision algorithms, to be applicable to human vision, should show the same types of tolerances.

But with regard to the use of camera lens in movies, it becomes quite clear why long lenses–those that are telephoto and nearly telephoto–are so popular and useful. First, and known for nearly a century, standard lenses tend to make people look like they have bulbous noses. Second, and corroborated by my results, long lenses provide a more nearly parallel projection of objects, and the distortions seen in these objects when a viewer looks at a slanted screen are significantly diminished. This enhances their efficacy considerably, despite the fact that it introduces the nonnatural situation of collapsing the apparent depth of a scene.

# REFERENCES

Cutting, James E. Perception With an Eye for Motion. MIT Press/Bradford Books, 1986(a).

Cutting, James E. The Shape and Psychophysics of Cinematic Space. Behav. Res. Meth. Inst. & Computers, vol. 18, 1986(b), pp. 551-558.

Cutting, James E. Rigidity in Cinema Seen From the Front Row, Side Aisle. J. Exp. Psychol.: Human Perc. & Perf., vol. 13, 1987, pp. 323-334.

Hagen, Margaret A.; and Elliott, Harry B. An Investigation of the Relationship Between Viewing Conditions and Preference for True and Modified Perspective With Adults. J. Exp. Psychol.: Human Perc. & Perf., vol. 4, 1976, pp. 479-490.

Hochberg, Julian. Representation of Motion and Space in Video and Cinematic Displays. In K.R. Boff, L. Kaufman, & J.P. Thomas (eds.), Handbook of Perception and Human Performance, Vol. 1, pp. 22:1-63. John Wiley, 1986.

Kubovy, Michael. The Psychology of Linear Perspective and Renaissance Art. Cambridge University Press, 1986.

Pirenne, M. H. Optics, Painting, & Photography. Cambridge University Press, 1970.

Shepard, Roger N. & Cooper, Lynn A. Mental Images and Their Transformation. MIT Press/Bradford Books, 1982.
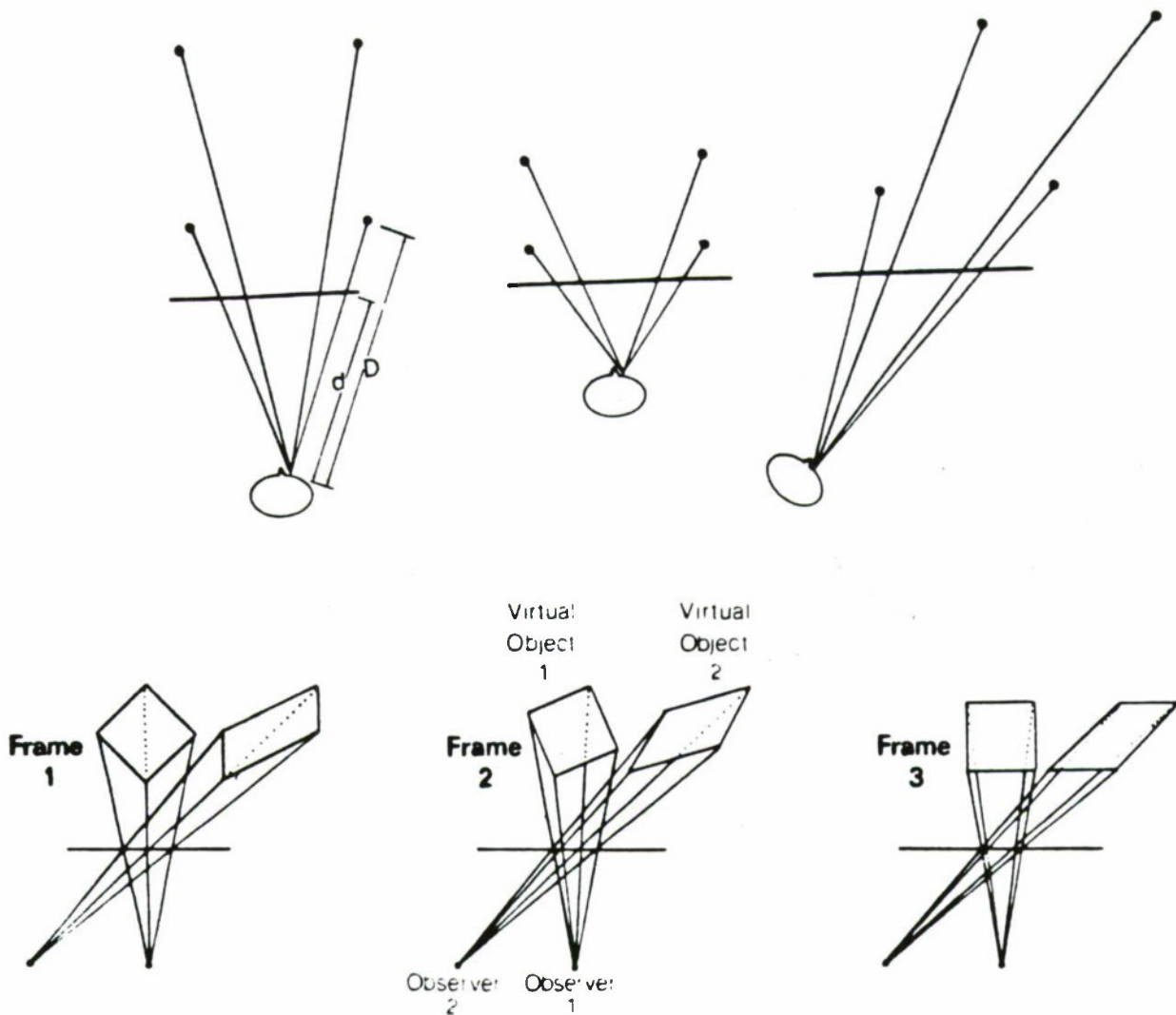
Figure 1. Reconstructive geometry and images. The upper panels show the reconstruction of four pillars in depth. Consider the left-most panel a representation of the real depth relations projected onto the image plane. If that plane is now a photograph, the pillars are fixed in position on the image plane. Thus, when an observer moves toward the plane, depth must be compressed, as in the upper middle panel. When the viewer moves to the side, all pillars slide over by differing amounts. The bottom panels show reconstructions of a moving square across three frames, from two viewpoints. Notice that the reconstruction for Observer 1 is rigid, but that that for Observer 2 is not (from Cutting, 1986a).
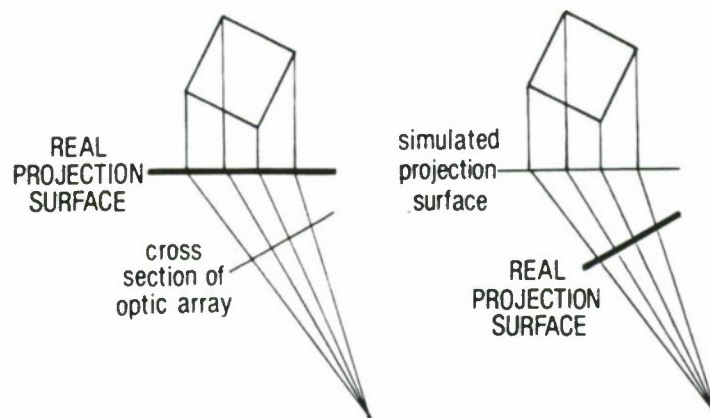
Figure 2. Arrangements of real and simulated projection surfaces that can remove image information from objects projected onto slanted screens (from Cutting, 1987).
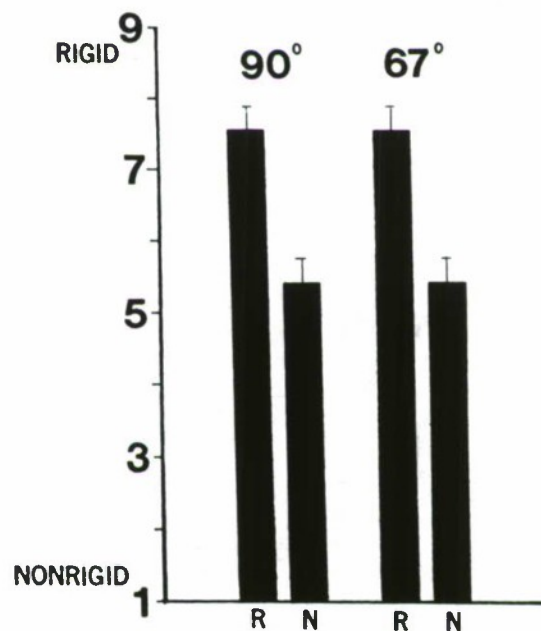


Figure 3. Selected results from Experiment 1. 90° and 67° are the two viewing conditions of interest, where 67° is the simulated screen slant as indicated in figure 2. R = rigid stimuli, N = nonrigid stimuli.
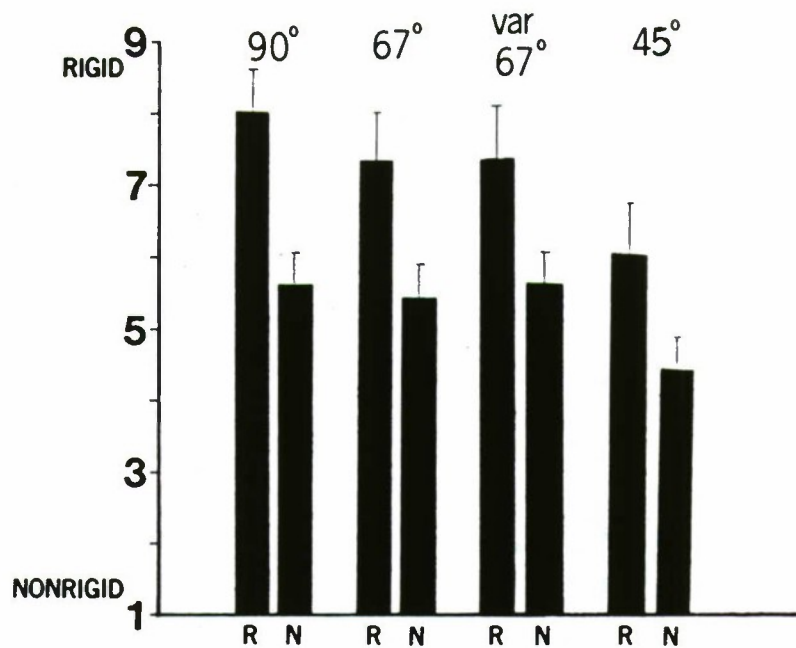
Figure 4. Selected results from Experiment 2. The added conditions are simulated screen slants of 45°, and one of variable slant (between 80° and 55°), averaging 67°. R = rigid stimuli, N = nonrigid stimuli.
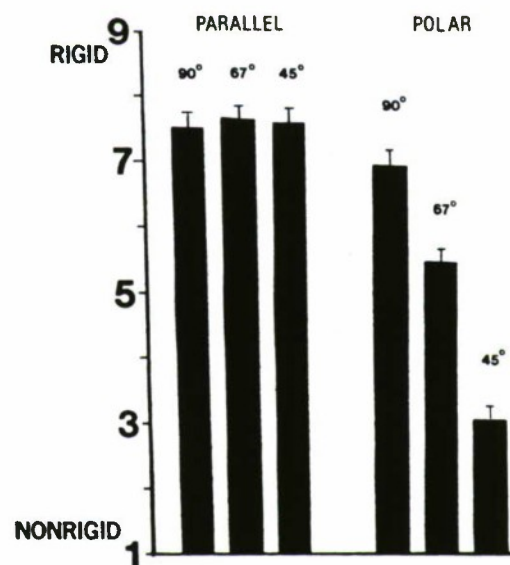


Figure 5. Another description of the results of Experiment 2, parsed according to projection.

# THE PERCEPTION OF GEOMETRICAL STRUCTURE FROM CONGRUENCE[1]

Joseph S. Lappin
Vanderbilt University
Nashville, Tennessee

and

Thomas D. Wason
Allotech Inc.
Raleigh, North Carolina

## 1. INTRODUCTION

The principal function of vision is to measure the environment. As demonstrated by the coordination of motor actions with the positions and trajectories of moving objects in cluttered environments and by the rapid recognition of solid objects in varying contexts from changing perspectives, vision provides real-time information about the geometrical structure and location of environmental objects and events.

Information about the geometrical structure of scenes, objects, and motions may be visually acquired not only by the exploration of natural environments, but also from artificial, human-designed displays. Photographs, drawings, movies, computer graphics, and other such artificial 2-D displays are widely and effectively used tools for communicating information about spatial structures. Understanding the basis for the effectiveness of such tools poses a special theoretical challenge, because the trigonometric mapping from the 3-D structures and motions portrayed in these displays to the optical patterns on the observer's retinae differs from the perspective projections that normally hold for vision in natural environments. Cutting (1987) has recently discussed the theoretical difficulties posed by this discrepancy between the projective geometry of movies versus that of natural vision, and he has also provided experimental demonstrations of the abilities of humans to perceive 3-D structure in movies viewed "from the front row side aisle."

The purpose of this paper is to examine the geometric information provided by 2-D spatial displays. We propose that the geometry of this information is best understood not within the traditional framework of perspective trigonometry, but in terms of the structure of qualitative relations defined by *congruences* among *intrinsic* geometric relations in images of *surfaces*. The mathematical details of this theory of the geometry of vision are presented elsewhere (Lappin, in press); the present paper outlines the basic concepts of this geometrical theory.

Traditionally, the structure of space–both the 3-D space of the environment and the 2-D space of the image–has been regarded as defined *a priori,* independently of the objects and motions contained within it. Indeed, the geometric structure of objects and motions is typically described by reference to extrinsic standards that define parallel and perpendicular directions and quantify relative magnitudes of distance extrinsic to the objects themselves.

When described in terms of this extrinsic framework, however, the geometry of vision is quite complicated: Metric[2] relations in the 2-D image plane cannot be isomorphic with metric relations in the 3-D environment; the perspective projection from 3-D spatial structures in the environment onto the 2-D image plane does not have a well-defined inverse. Therefore, the recovery of information about the geometric structures and locations of the environmental objects has often been thought to require supplementary information about the perspective position of the observer or about the structure and location of the objects. The 2-D optical images alone have seemed insufficient.

But the assumption that vision begins with an abstract structure of space as a prior standard for describing environmental objects begs the question. The basic problem of vision is to find a measurement structure for representing the spatial characteristics of observed scenes, objects, and events. Such a measurement structure is generally not given beforehand, but must be discovered in the organization of the empirical observations themselves.

## 2. INTRINSIC GEOMETRY OF SURFACES AND IMAGES

When described in terms of the intrinsic geometry of *surfaces,* the geometry of vision becomes much simpler. In the first place, the mapping of a visible region of an environmental surface onto its optical image is a mapping from one 2-D manifold onto another. The derivatives and singularities of the surface–its slopes, peaks and valleys, inflections, saddlepoints, and occluding edges–are isomorphic with the derivatives and singularities of the image. This is true for images described by gradients of texture, motion parallax, or stereoscopic disparity (Koenderink and van Doorn, 1975, 1976a,b,c, 1977). Although the isomorphism does not hold for images described by luminance gradients, partly because of the additional influence of the direction of illumination, it is still true that the intrinsic surface structure (in particular, the parabolic lines, which are inflections of curvature that separate regions of convexity and concavity) is systematically related to the differential structure of the image (Koenderink and van Doorn, 1980). Because the differential structures of the two manifolds are essentially isomorphic with one another, the ordinal topography of the visible region of an environmental surface is fully described and recoverable by its optical image.

Furthermore, the specific mapping between curves and forms on the environmental surface and their corresponding images on an observer's retina may be locally described simply by a *linear*

---

[2]The term metric is used in a conventional mathematical sense, referring in this context to measures of distance over a potentially curved surface. A relation $m(a,b)$ between two elements a and b is said to be a metric relation if it satisfies the following axioms for all elements a, b, and c:

     positivity: $m(a,b) \geq 0$
     symmetry: $m(a,b) = m(b,a)$
     reflexivity: $m(a,a) = 0$
     triangle inequality: $m(a,c) \leq m(a,b) + m(b,c)$.

*coordinate transformation* between the derivatives on the two manifolds. This linear approximation holds for "infinitely small" surface patches that may be locally approximated by a tangent plane at that location. This linear mapping of the surface onto its image also has a well-defined inverse. Accordingly, the local structure of the surface may be obtained from the local structure of its image by a linear coordinate transformation.

These simple relationships between the surface and its image involve the *derivatives* on the two manifolds. The linear transformation that best describes the relationship between these two manifolds at any given point is given by the partial derivatives of the two coordinate systems. Thus, if $O^2$ represents the 2-D manifold of the object surface, and if $R^2$ represents the 2-D manifold corresponding to the observer's retina, then the linear differential map $v: O^2 \rightarrow R^2$ is specified by the following Jacobian matrix of partial derivatives:

$$V = \begin{bmatrix} \partial r^1/\partial o^1 & \partial r^1/\partial o^2 \\ \partial r^2/\partial o^1 & \partial r^2/\partial o^2 \end{bmatrix}$$

Suppose that $[dO] = [do^1, do^2]$ is a 2x1 column vector that specifies an infinitesimal displacement on the surface in terms of two intrinsic coordinates on the object surface, and suppose that $[dR] = [dr^1, dr^2]$ is a corresponding description of the image of this vector in terms of the intrinsic coordinates of the retina. Then the transformation between these two coordinate systems produced by the optical projection from the object to its image on the retina is given by the linear equation

$$[dR] = V[dO]$$

and the inverse map is given by

$$[dO] = V^{-1}[dR]$$

where $V$ is the Jacobian matrix given above. (The form of this equation is independent of the specific coordinate systems used to specify positions on the two manifolds. The coordinates need not intersect at right angles nor even be straight lines; they need only be differentiable and to provide a unique specification of each position on the manifold. The generality of this representation seems especially relevant to vision, where no specific coordinate system can be assumed beforehand for any particular environmental surface.[3] ) The important point is that *the local structure of the retinal image of a given surface is described by this Jacobian matrix of partial derivatives, V*. The entries in this matrix vary as a function of position on the surface, with variations in the values of these entries reflecting variations in the orientation and curvature of the surface.

The same approach can also be used to describe the relationships with a third 2-D manifold associated, for example, with an intervening display image such as a movie or photograph. Suppose that $I^2$ represents the manifold of such an intervening image, that $a: O^2 \rightarrow I^2$ represents the

---

[3]For concreteness, we may assume that the coordinates reflect the spatial arrangement of the gradients and singularities of the surface—e.g., tending to run parallel and perpendicular to the gradients of curvature of the surface and to the boundary contours, corners, and parabolic lines (which separate structurally distinct regions). We need not assume that these coordinates have specific numerical values, only that they are differentiable and uniquely label every location on the surface.

differential map between these two manifolds, and that b: $I^2 \rightarrow R^2$ is the visual map from the display image onto the retinal manifold. Then, using the chain rule, the two successive maps can be combined by a composition of the two functions, $v = (b \cdot a): O^2 \rightarrow R^2$. Similarly, the coordinate transformation corresponding to this chain would be given by a linear equation of the following form:

$$[dR] = BA[d0],$$

where the matrix product $BA = V$ again provides a linear coordinate transformation functionally equivalent to the previous construction.

Representation of the *metric* structure of the surface requires an embedding of the 2-D manifold of the surface or its image into the 3-D manifold of Euclidean 3-space, $E^3$. Suppose that $[dX] = [dx^1, dx^2, dx^3]$ is a 3x1 column vector giving the three orthogonal cartesian coordinates of an infinitesimal displacement on the object surface. Then the perspective coordinate embedding of the image of the surface into $E^3$, p: $R^2 \rightarrow E^3$, is given by a linear coordinate transformation of the following form:

$$[dX] = PV[d0]$$

where P is a 3x2 matrix of partial derivatives, $P = [\partial x^k/\partial r^i]$, with $k = 1,2,3$ and $i = 1,2$. Measures of metric relations require a quadratic expression similar to the Pythagorean formula for distance in $E^3$. The metric tensor that provides the measure of distance on the surface is obtained by substituting from the above equation for the vector $[dX]$ in the Pythagorean formula:

$$
\begin{aligned}
ds^2 &= [dX]^T [dX] \\
&= [PV[d0]]^T PV[d0] \\
&= [d0]^T V^T P^T PV[d0] \\
&= [d0]^T V^T P^* V[d0],
\end{aligned}
$$

where $P^* = P^T P$ is a symmetric 2x2 matrix with quadratic entries of the form

$$P^* = [\sum_k (\partial x^k/\partial r^i)(\partial x^k/\partial r^j)].$$

Thus, the entries in this matrix provide a measure of squared distance on the object surface at a particular position on the retina corresponding to the image of the surface. The length of any arbitrary curve on the surface is obtained by integrating the quantities ds defined in the preceding equation at each position along the curve.

The three independent parameters of the matrix $P^*$ are not given directly by a single stationary image of an isolated local surface patch. In certain special cases these perspective parameters and therefore the metric structure of the local surface patch are determined, up to a scalar, simply by the motion of the local patch. More generally, however, these perspective parameters must be derived from more global constraints on the image structure associated with the observer's position

and motion within the 3-D environment. In general, the perspective embedding of the image into $E^3$ is revealed by actual or implied motions of objects within the space.

## 3. METRIC STRUCTURE FROM CONGRUENCE

Although geometric relations are often described in terms of extrinsic coordinate systems in which directions and distances are defined *a priori,* it is important in many applications to derive the structure of space from more fundamental qualitative relationships among the objects and events contained within it. This was the case, for example, in the development of relativistic physics, where the symmetries of observations associated with the velocity of light and with gravitation were used to construct spaces in which the lawful relations among observed variables could be expressed in simpler and more general form (Einstein *et al.* 1923, 1952; Misner, Thorne, and Wheeler, 1973). The same strategy has also been employed in formulating the theoretical foundations of measurement (Krantz *et al.*, 1971; Luce, 1978; Luce and Narens, 1983). That is, symmetries of qualitative relations under various physical operations and under varying conditions of observation may often be used as a foundation for quantitative equations that describe empirical laws of nature.

Analogously, the geometry of vision may also rest upon the symmetries of intrinsic qualitative relations in the spatio-temporal optical images rather than on the prior metric structure of an extrinsic coordinate system. Because metric relations in the 3-D environment are not isomorphic with those in the 2-D image, and because the optical projections of environmental objects onto the retinae change with the perspective positions of the displays and observers, the extrinsic framework of space is neither constant nor readily accessible to vision. Instead, we hypothesize, the metric structure of environmental objects and spaces may be induced from the isometries of moving objects.

This conception of the geometry of vision is a continuation of ideas developed by Gibson (1950, 1957) about the importance of the concepts of invariance and transformations for perception (Lombardo, 1987). Gibson's (1950) conception of the visual information provided by such "higher order variables" as a texture density gradient was based on the idea that gradients of repeated structural relations specified the projective transformation of a surface onto an image and also specified an intrinsic scaling of the 3-D space in which the surface texture was homogeneous. The same conception was subsequently expanded (e.g., Gibson, 1957) to emphasize the information provided by the continuous transformations of optical flow produced by moving objects and moving observers. These deformations of the optical images were believed to enable the perception of both the structural invariants and the projective transformations associated with the motions of objects and observers in 3-D space.

The essential ideas underlying this conception of geometry were described by the mathematician Killing (1892)[4] :

> Every object covers a space at every time. The space covered by one object cannot simultaneously be covered by another object.

---

[4]We are grateful to Jan Koenderink for bringing this paper to our attention and to Bernd Rossa for translating the paper from the original German.

Every object can be moved. If an object covers the space of a second object at any time, then the first object can cover any space covered by the second object at any (other) time.

Every space (object) can be partitioned. Each part of a space (object) is again a space. If A is a part of B and B is a part of C, then A is a part of C, where A, B, and C may be either spaces or objects. [p. 128]

These three principles, which are the first of eight principles from which Killing derives a general theory of geometry, provide qualitative criteria for defining the equality or congruence of spaces and objects: Two spaces are congruent if and only if they can be covered by the same object. Two objects are congruent if and only if they can cover the same space. Thus, objects and spaces constitute mutually interdependent relational structures. The metric structure of both may be derived from elementary qualitative properties of *differentiability* and *congruence under motion*. (By definition, "motions" are isometric transformation groups.) (Also see Weyl, 1952, and Guggenheimer, 1963, Sect. 11-2.)

This conception of form and space provides a basis for understanding how visual information about the metric structure and dimensionality of objects and spaces may be gained from "motions" or transformations which bring objects at one position in space into congruence with those at other positions. The metric equality of neighboring spaces successively occupied by the same object and the equality of separate parts of an object which successively occupy the same space may be determined from the motions of objects. Accordingly, the dimensionality of visible spaces and objects need not be restricted to the two coordinate dimensions of the image. Rather, the dimensionality may be associated with the number of parameters needed to bring an object at one location in space into congruence with an object at another location.

In certain special cases the metric structure of a given surface patch may be *locally* determined (up to a scalar) by its moving images, independent of global properties of the retinal image as a whole. If the *trajectory* of the moving patch is also a *surface in space-time* with constant curvature equal to that of the object patch, then of course the metric tensor for this spatio-temporal surface remains constant over the surface. Motion of the object patch from one region of the spatio-temporal surface to another does not change the mapping of the surface onto the retina, and the contravariant tensor coefficients for this projective mapping of the object patch and its trajectory onto the retina vary only as one-parameter functions of time. Accordingly, the perspective parameters for embedding the retinal images of this surface into $E^3$ also vary as one-parameter functions of time or of retinal position (which are correlated in this case). The simplicity of these relationships between the differential structure of the object surface, its trajectory in space-time, and the retinal images of these surfaces involves sufficiently few unknown perspective parameters that these are determined by the invariance of the metric tensor of the surface patch under motion. That is, suppose that $V_o$ and $P_o$ are the Jacobian matrices for the visual and perspective coordinate transformations, respectively, for an initial retinal image of the surface patch, and suppose that $V_t$ and $P_t$ are the corresponding matrices for a second retinal image of the same surface patch following a one-parameter motion onto another position along its constant-curvature trajectory. The equivalence of the geometric structure of the two retinal images can be expressed by the equation

$$V_o^T P_o^* V_o = V_t^T P_t^* V_t$$

where $P_t = m(P_0)$ is a one-parameter transformation of $P_0$. This matrix equation involves four independent linear equations in four unknowns–the three independent perspective parameters of $P_0$ and the transformation of these by the parameter t.

Specific examples of this special case include a sphere that rotates around an axis (different from the direction of gaze) through its center (e.g., Lappin, Doner, and Kottas, 1980; Doner, Lappin, and Perfetto, 1984) and planar patterns that rotate within the same plane (Lappin and Fuqua, 1983) tilted with respect to the retinal image. In both of these cases the time-varying set of positions of the surface patch form a surface of revolution in space-time generated by a one-parameter transformation group (the magnitude of the rotation). In general, the metric tensor for the images of the moving surface patch remains invariant under the motion (i.e., its Lie derivative is zero) if and only if the vector field of this group of isometries (the "Killing vector") is a one-parameter group that generates a *surface of revolution* (Guggenheimer, 1963, pp. 272-273). Thus, because the moving object forms a surface whose images are generated by a one-parameter transformation, the perspective parameters for embedding this spatio-temporal surface into $E^3$ are determined up to a scalar by the invariant metric structure of the given surface patch. Indeed, the experimental results of Lappin, Doner, and Kottas (1980) and Doner, Lappin, and Perfetto (1984)–for the perceived shape of a random-dot sphere rotating about a vertical axis through its center–and of Lappin Fuqua (1983) for the perceived inter-point distances among three collinear points rotating in a plane–demonstrated just this invariance of visually perceived metric structure under motion even though the optical displays contained unnaturally exaggerated amounts of polar projection.

In general, however, the metric structure of moving objects cannot be recovered from only local properties of their retinal images. Instead, the perspective parameters of the projection from $E^3$ onto the retina must be recovered more global constraints on the images. Perspective projection from $E^3$ onto a plane produces a *hyperbolic geometry* in the plane, where mutually parallel lines converge toward a common vanishing point and all sets of parallel lines converge toward a common horizon line. The position of this horizon line in the visual field is equal to the observer's eye-height. Accordingly, all lines parallel to the observer's motion through the 3-D environment converge toward a common vanishing point on the horizon that specifies the observer's momentary position and trajectory through the visible environment. The images of such parallel lines in $E^3$ are generated by the retinal image trajectories of features of stationary environmental objects as the observer moves through the environment. Thus, the location of this horizon line and of such vanishing points constitute parameters that characterize the given hyperbolic space and its relation to $E^3$. Like Euclidean space, hyperbolic space is also characterized by congruence and isometry of form under motion. Thus, congruence relations among visible forms must specify this global perspective embedding of the retinal image into $E^3$. Although we have not yet completed the mathematical analysis of this situation, the following illustrations may help to convey the rationale for this conception of the geometry of vision.

## 4. CONGRUENCES IN IMAGES

The potential for constructing spaces from congruences among imaged forms has been wonderfully illustrated by M. C. Escher. For example, he has often used translational symmmetry of a replicated form to define a 2-D plane. Both the metric structure of this space and also its 3-D orientation parallel to the image plane are specified by the translational symmetry. The elementary

component form is also defined by its recursion in the image rather than by the familiarity of the form itself.

Symmetries in 3-D Euclidean space are exhibited in figure 1, where the congruence of swan-like component forms is obtained by translations and rotations in a 3-D space. The 3-D metric structure of the space is implied by the congruence of the recurring forms in separate regions of the space. The perspective mapping of this space onto the 2-D image plane is also induced by this congruence of the component forms. Thus, the perspective trigonometry is derived from the congruence; the fundamental property is the congruence rather than the trigonometry.

In the preceding example, congruence is defined among stationary and concurrent forms. The "motion" that brings a form in one location into congruence with a form in another location is abstract, rather than an actual trajectory in space-time. If one generalizes the concept of an image from a stationary 2-D spatial array to a space-time volume in which the spatial structures are extended in time, then the same principle of congruence illustrated in Escher's art can be applied to the specification of spaces by the motions of single forms.

The schematic diagram in figure 2 illustrates three conceptually different types of congruence in images. Figure 2B is like that in the Escher print, where the image is a stationary 2-D pattern in which a single cube-like structure is recursively positioned at a sequence of neighboring spatial positions. The 3-dimensionality of the space is induced by the continuous linear change in the 2-D lengths of the contours of the cube as a function of its position in the image plane. This linear relation between 2-D length and position corresponds to a particular perspective mapping of 3-D space onto the image plane. Thus, the continuous linear relation among neighboring regions of the image of a single connected surface specifies the perspective mapping of a 3-D space onto the 2-D image.

In figure 2A the same perspective mapping is defined by a temporal sequence of spatial images as the cube is translated through space from position $P_1$ to position $P_n$. The linear transformation that corresponds to the perspective projection of a plane slanted in depth is now specified by a function in space-time, though the geometric relation between the image and the depicted space obviously is essentially the same as in figure 3B. In both cases, relationships among neighboring image regions correspond to relationships among neighboring regions of a smooth surface. The perspective relation between the image and the 3-D space in which the surfaces, objects, and motions reside is specified by the linear relationship between the lengths of the contours and their positions in the image.

Figure 2C illustrates a slightly different case in which the structure of a space is specified by congruences among simultaneous *motions* of separate forms at separate locations in the image, as if the forms were connected and moved in 3-D space. This situation might be produced, for example, by motions of the observer or image plane (e.g., a movie or video camera) within a 3-D environment. In this example two cubes, at positions $P_1$ and $P_n$ in 3-D space, are simultaneously displaced in a sequence of four successive translations. The perspective mapping from the 3-D space in which these events occur onto the 2-D image of the events may be specified by the functional relation between the magnitudes of the velocities and their locations in the image plane. Although the forms at positions $P_1$ and $P_n$ in this particular illustration are both cubes that are potentially congruent under the same transformations that would bring the motions of the two cubes into congruence, this spatial congruence is not necessary and provides in this case an addi-

tional redundant specification of the perspective transformation of the 3-D space onto the 2-D image.

The geometric relation between the concurrent motions of just two forms as in figure 2C is not generally sufficient to specify the perspective transformation that has yielded the observed spatio-temporal image. By the fundamental theorem of plane perspectivity (Delone, 1963), the perspective mapping of four points in general position (where no three points are collinear) in one image plane onto a corresponding set of four points in another image plane is necessary and sufficient to ensure that all of the remaining points are in isometric correspondence in the two planes. Thus, for a set of four or more points in a single plane, the concurrent motions of the images of these points in another plane are in principle sufficient to specify the perspective transformation between these two planes and to specify the metric structure of the spatial relations within these planar images.

This geometric relationship endows spatial as well as moving images with considerable capacity for carrying information about the geometric structure of the environmental surfaces depicted in the images: The geometric structure of an infinitesimally small patch on any arbitrarily curved but smooth surface may be locally approximated by a tangent plane at that location, and the perspective mapping of this tangent plane onto an image plane may be described by a linear coordinate transformation. The parameters of this linear transformation vary with the relative 3-D orientation (the direction of tilt and the magnitude of slant) and distance of the environmental surface in relation to the image plane. The perspective parameters which embed the image of the surface into $E^3$ and thereby determine the metric structure of the surface are those parameters that will yield the self-congruence of the same object at different locations within the depicted scene.

## 5. EXPERIMENTAL EVIDENCE

In addition to the evidence provided by the illustrations, by everyday visual experience in viewing both natural environments and artificial spatial displays, and by the capabilities of moving observers to coordinate their actions with the identities, positions, and trajectories of environmental objects, the hypothesis that perceived geometric structure derives from the congruences of moving and movable objects is also supported by experimental evidence. A vast amount of experimental evidence appears consistent with this hypothesis, but we mention here only a few experiments that seem to provide more direct support for this hypothesis.

One of the relevant investigations is that of Cutting (1987). Judgments of the apparent rigidity of rotating rectangular solids were evaluated in a variety of experimental display conditions, including both rigidly and nonrigidly rotated figures and displays that simulated varying degrees of polar versus parallel projection, and varying degrees of slant of the projection screen relative to the direction of the perspective convergence point. He found good discrimination of rigid versus nonrigid figures in displays with approximately parallel projection, essentially independent of the degree of simulated screen slant (90°, 67°, 45°, or varying between 80° and 55°), even when the simulated slant was varied sinusoidally during a given trial. Although the figures appeared to move nonrigidly in conditions with polar projection onto screens slanted at 45°, the results generally demonstrated the robustness of perceived structural rigidity under at least moderate screen slants and moderate viewing distances. These results challenge many conventional assumptions about the geometrical information for perceiving the spatial structure of form. Cutting concludes

that these results probably reflect the insensitivity of vision to the distortions produced by optical projections, but this interpretation rests upon assumptions about the definition of visual space by the metric structure of 2-D display screens and retinae. An alternative interpretation is that vision is very sensitive to spatial relations defined in another way—by the congruence of form under perspective transformations.

Evidence that vision is indeed very sensitive to the spatial structure of moving forms and that this structure is associated with invariant spatial relations in depth rather than the projected 2-D positions is provided by experiments reported previously by Lappin and Fuqua (1983). They evaluated observers' acuities in detecting a displacement (a stationary offset in 3-D space) of a point from the 3-D center of an imaginary line segment defined by moving patterns of three collinear points. The points were rotated in computer-controlled CRT displays as if around an axis slanted in depth by amounts varying between trials from 0° (no slant) to 60°. Very small displacements were accurately detected—displacements greater than 1% of the 3D distance between the two outer points could be detected above chance, and displacements of 4% were detected at approximately 90% accuracy. The essential 3-dimensionality of the perceived spatial relations was demonstrated by the following findings: (1) Detection accuracy was independent of either the magnitude or variability of the slant of the axis of rotation in depth. (2) Distance-like measures of the detection accuracy (similar to the signal detectability measure $\underline{d}'$) were linearly related to the physical distance of the displacement in 3-D space, with discriminability being proportional to physical displacement distances above about 1%. (3) The accuracy for detecting any given displacement was the same in displays with parallel and with polar perspective, although in the latter displays points centered in 3-D depth were not centered in the projected 2-D images. The differences in spatial positions between the parallel and polar displays were visually resolvable, however. (4) When the task required detection of displacements from the projected *2-D* centers of the line segments in displays with polar projections, accuracies were not significantly above chance. The subjective appearance of the latter displays was that the three points were still seen as rotating in depth, but the middle point appeared neither centered nor rigidly attached to the two outer points.

Thus, these findings suggest that vision may often be unaffected by the 2-D optical "distortions" in cinema not merely because these spatial differences cannot be resolved by vision, but because they do not constitute the geometrical information for perceiving the spatial structure of moving patterns. Apparently, perceived spatial structure derives from congruences of form under perspective transformations.

Evidence about the role of such congruences in *stereoscopic* form perception has been provided by recent experiments described by Lappin (in preparation). The purpose of these experiments was to determine whether the stereoscopic perception of 3-D structure might be shaped by the congruences of form associated with motion in depth, rather than by the binocular disparities as such. The experiments were motivated by the theoretically challenging fact that for any given magnitude of binocular disparity between the horizontal separations of a pair of points in each eye, the associated separation in depth increases rapidly and nonlinearly with the viewing distance from the observer to the points in question: How then is the stereoscopic perception of form and depth calibrated for variations in viewing distance? Does this require "interpretations" of retinal disparities based on extra-retinal information about the viewing distance? Alternatively, might the perceived geometric structure of surfaces in depth be based on the invariance of the intrinsic geometric structure of the surface under the perspective transformations associated with stereoscopic disparities and with motions in depth? The theoretical problem is related to those in understanding the apparent "paradoxes" of cinema.

In one of these experiments, observers were presented with two very slightly different ellipses, in which the vertical axis was either 3% greater or less than the length of the horizontal axis. These ellipses were displayed as if in a plane slanted in depth by either 50° or 60° varying randomly from one trial to the next. Thus, the projected forms were always elliptical, depending on the magnitude of the slant as well as on the shape of the ellipse as measured in its own plane in depth. Stereoscopic information about the shapes and slants of these patterns was also manipulated by random variations in the magnitude of the disparities with which the forms were displayed, using disparities that were appropriate for either one-half or one-quarter of the actual viewing distance at which the patterns were seen. Thus, there were eight alternative stimulus patterns which randomly varied between trials.

There were four main experimental conditions—in which the forms were either rotated in depth or were stationary, and in which the experimental task was either *shape-discrimination* between the two alternative ellipses or *disparity-discrimination* between the two alternative disparity values. If stereoscopic information about 3-D structure is scaled by the congruences of moving forms, then shape discrimination should be accurate when the forms were rotated in depth, independently of the distortions and variability produced by the exaggerated binocular disparities. Indeed, this is just what happened: Shape discriminations were very accurate when the forms were moving, and were uncorrelated with the variations in either slant or disparity. Not surprisingly, shape discriminations were near chance accuracy when the forms were stationary because of the perceptually inseparable conjoint effects of variations in slant and disparity. For the disparity-discrimination task, however, motion had the opposite effects: Discriminations between the two alternative disparity values were more accurate for the stationary than for the moving forms, evidently because the congruence of the moving forms tended to obscure differences between the stationary disparity spaces.

Thus, these results indicate that the visual scaling of 3-D structure from stereoscopic disparity derives from the congruences of the perspectively changing forms. Analogous to the case for stationary pictures and optic flow patterns, binocular disparity per se may have only an indirect relation to the perceived depths.

# REFERENCES

Cutting, J. E. (1987) Rigidity in cinema seen from the front row, side aisle. J. Exp. Psychol: Human Percept. Perform. 13, 323-334.

Delone, B. N. (1963) Analytic geometry. In A. D. Aleksandrov, A. N. Kolmogorov, and M. A. Lavrent'ev (Eds.) Mathematics: Its content, methods, and meaning. Cambridge: MIT Press.

Doner, J., Lappin, J. S., and Perfetto, G. (1984) Detection of three-dimensional structure in moving optical patterns. J. Exp. Psychol: Human Percept. Perform., 10, 1-11.

Einstein, A., Lorentz, H. A., Weyl, H., and Minkowski, H. (1923/1952) The principle of relativity. New York: Dover.

Gibson, J. J. (1950) The perception of the visual world. Boston: Houghton-Mifflin.

Gibson, J. J. (1957) Optical motions and transformations as stimuli for visual perception. Psychol. Rev., 64, 288-295.

Guggenheimer, H. W. (1963) Differential geometry. New York: McGraw-Hill.

Killing, W. (1892) Ueber die Grundlagen der Geometrie. J. fur die reine und angewandt Mathematik, 109, 121-186.

Koenderink, J. J. and van Doorn, A. J. (1975) Invariant properties of the motion parallax field due to the movement of rigid bodies relative to the observer. Opt. Acta, 22, 773-791.

Koenderink, J. J. and van Doorn, A. J. (1976a) Local structure of movement parallax of the plane. J. Opt. Soc. Amer., 66, 717-723.

Koenderink, J. J. and van Doorn, A. J. (1976b) Geometry of binocular vision and a model for stereopsis. Biol. Cyber., 21, 29-35.

Koenderink, J. J. and van Doorn, A. J. (1976c) The singularities of the visual mapping. Biol. Cyber., 24, 51-59.

Koenderink, J. J. and van Doorn, A. J. (1977) How an ambulant observer can construct a model of the environment from the geometrical structure of the visual inflow. In G. Hauske and E. Butenandt (Eds.) Kybernetik. Munich: Oldenberg.

Koenderink, J. J. and van Doorn, A. J. (1980) Photometric invariants related to solid shape. Opt. Acta, 27, 981-996.

Krantz, D. H., Luce, R. D., Suppes, P., and Tversky, A. (1971) Foundations of measurement. New York: Academic Press.

Lappin, J. S. (in press)  Perceiving the metric structure of environmental objects from motion and stereopsis.  In R. Warren and A. H. Wertheim (Eds.) The perception and control of self-motion.  Hillsdale, N.J.:  Lawrence Erlbaum.

Lappin, J. S. (in preparation)  The scaling of stereoscopic form by congruence under motion.

Lappin, J. S., Doner, J. F., and Kottas, B. L. (1980)  Minimal conditions for the visual detection of structure and motion in three dimensions.  Sciences, 209, 717-719.

Lappin, J. S. and Fuqua, M. A. (1983)  Accurate visual measurement of three-dimensional moving patterns.  Science, 221, 480-482.

Lombardo, T. J. (1987)  The reciprocity of perceiver and environment:  The evolution of James J. Gibson's ecological psychology.  Hillsdale, N.J.:  Lawrence Erlbaum.

Luce, R. D. (1978)  Dimensionally invariant numerical laws correspond to meaningful numerical relations.  Phil. Sci., 45, 1-16.

Luce, R. D. and Narens, L. (1983)  Symmetry, scale types, and generalizations of classical physical measurements.  J. Math. Psychol., 27, 44-85.

Misner, C. W., Thorne, K. S., and Wheeler, J. A. (1973)  Gravitation.  San Francisco:  Freeman.

Weyl, H. (1952)  Space-time-matter.  New York:  Dover.
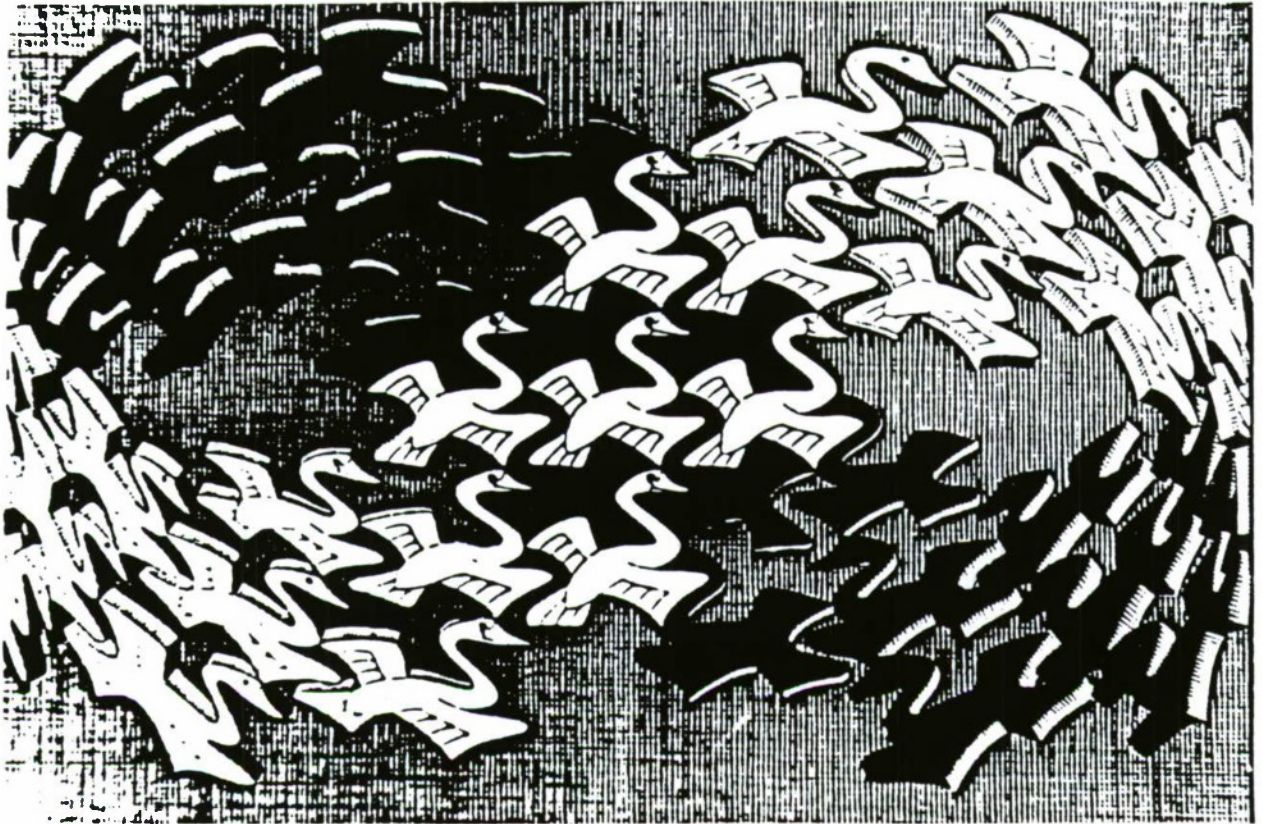
Figure 1.– "Swans," etching by M. C. Escher, 1956. © 1988, M. C. Escher heirs/Cordon Art-Baarn-Holland.
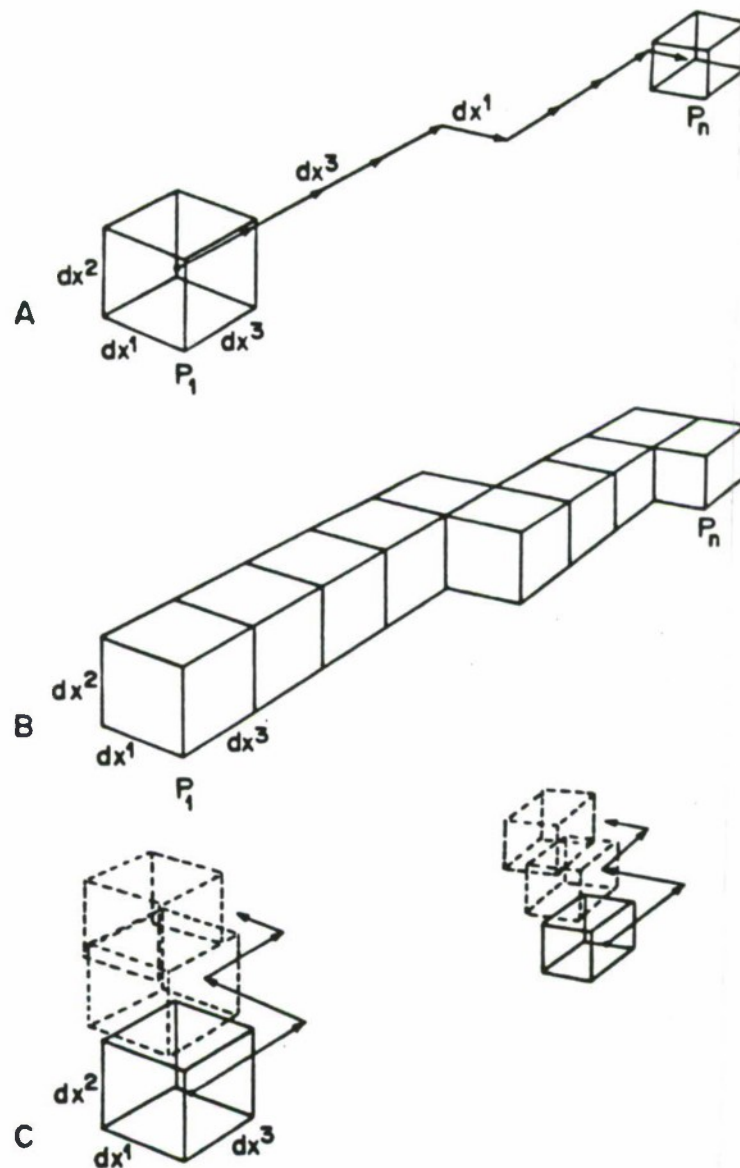
Figure 2.– Three types of congruences in images. (A) A cube in position $P_1$ is moved in a temporal sequence of displacements through 3-D space to position $P_n$. A single object appears in a trajectory through space-time. (B) The same cubic form as in A appears simultaneously in positions $P_1$ and $P_n$, connected in this case by a spatial series of cubes. A 3-D space is defined by the congruences of the spatial series of repeated component forms. (C) Two objects are moved concurrently by a sequence of displacements as if rigidly connected. The 3-D structure of the space is indicated in this case by the congruence of the motions in the separate spatial regions rather than by the congruences of the spatial forms as in the other two panels.

# EXOCENTRIC DIRECTION JUDGEMENTS
# IN COMPUTER-GENERATED DISPLAYS
# AND ACTUAL SCENES

Stephen R. Ellis, Stephen Smith, Michael W. McGreevy,
and Arthur Grunwald
NASA Ames Research Center
Moffett Field, California

## INTRODUCTION

One of the most remarkable perceptual properties of common experience is that the perceived shapes of known objects are constant despite movements about them which transform their projections on our retina. This perceptual ability is one aspect of shape constancy (Thouless, 1931; Metzger, 1953; Borresen and Lichte, 1962). It requires that the viewer be able to sense and discount his or her relative position and orientation with respect to a viewed object. This discounting of relative position may be derived directly from the ranging information provided from stereopsis, from motion parallax, from vestibularly sensed rotation and translation, or from corollary information associated with voluntary movement.

The measurement of shape constancy usually involves requesting that the viewer make some estimate of the geometric properties of an object, such as the apex angle of a isosceles triangle. Significantly, shape constancy does not disappear during static, monocular viewing, but its basis under these conditions must be different, since sensed motion is not involved. In a static image, shape constancy amounts to the recognition that each of a variety of views of the objects in the scene are all views of the same objects. This perceived constancy may be based on consciously or unconsciously accessed information concerning alternative views of the objects. These "memories," however, need not be of complete objects, since perceived constancy may be based on recall of only some salient features, such as parallelism of significant planes of the object.

In the absence of information directly providing range and orientation, as when viewing realistic pictures, the viewer's relative position with respect to an object can be only indirectly inferred from the projection of the object itself and its surround. The information in the projected lines of sight in the optic array can be used to infer the relative position of the viewer only if the viewer has at least a partial internal 3D model of the viewed objects and their surround (Grunwald and Ellis, 1986; Wallach, 1985). Thus, "shape constancy" in static, monocular scenes is somewhat circular, since the necessary shape information required to infer relative viewing position is itself the shape of the object in question. Nevertheless, shape constancy can be obtained through an interactive process if the viewer has a variety of static views of the same scene or object from different viewing positions and is able to construct appropriate correct hypotheses regarding the shapes. Because of inherent regularities in the world, viewers are usually quite good at forming appropriate shape hypotheses in natural environments (Gregory, 1966). But they can be tricked (Ittelson, 1952; Hochberg, 1987).

Shape constancy may be generalized to constancy of interrelations among objects in a spatial layout. Just as the shape of an object ordinarily appears constant when a viewer moves with respect to it, so too do the spatial interrelations among objects generally appear constant during

corresponding movement of a viewer (Pirenne, 1970; Wallach, 1985; also see Ellis, Smith, and McGreevy,1987; Goldstein 1987). Piaget's decentering task, which requires that one imagine how a scene would appear from an external viewpoint, is an experimental scenario that particularly exercises this type of constancy (Piaget, 1932).

The Piaget decentering judgement is formally similar to that required of someone using a map to establish viewer orientation with respect to some exocentric landmark. When based on a map in which there is a marker representing the viewer's position, this judgement constitutes an exocentric direction judgement (Howard; 1982). In recent experiments we have examined a specific instance of this judgement by presenting subjects with computer-generated, perspective views of three-dimensional maps that have two small marker cubes on them (fig. 1). One marker represented the subject's assumed position on the map, i.e. his or her reference position. The other represented a target position. The subject's task was to make an exocentric direction judgement and estimate the relative azimuth of the target direction with respect to a reference direction parallel to one axis of the ground reference. In the previous experiments this reference was typically a full grid.

Interpretations of recent systematic measurements of these exocentric judgements have suggested that the observed patterns of error can be analytically described in terms of an external world coordinate system rather than a viewing coordinate system centered and aligned with the view direction. (McGreevy and Ellis, 1986; McGreevy, Ratslaff, and Ellis, 1985). In these experiments in which scenes were viewed from the center of projection direction, errors were observed in which the subjects exhibited a kind of equidistance tendency in that they judged the target cubes to be closer to the axis crossing the reference axis than they actually were. The same bias appeared independent of viewing direction, and thus the patterns of direction judgement error exhibited a kind of position constancy; that is, the errors were functions of the physical positions of the targets and not the subject's view of them.

Since the subjects were not allowed freedom to move the display's eye point during the individual judgements, position constancy would have to be based on assumed properties of the objects and features of the scene. The most likely feature that could provide the basis for this constancy is the ground reference meshed grid. Since the subjects may reasonably make the correct assumption that the grid axes are orthogonal, the grid can provide information about the compressive and expansive perspective effects due to the viewing parameters and allow the viewer to discount them. The information is provided most directly in the projected angle between the reference axis and the crossing axis. (Attnaeve and Frost, 1969; Ellis, Smith, and McGreevy, 1987).

Accordingly, removal of the crossing axis should remove the most direct information that allows the viewer to discount the geometric consequences of his or her particular viewing direction. Thus, a display used for the same kind of exocentric direction judgements, but lacking the crossing axis, should not exhibit position constancy. Direction judgement errors should now depend upon the viewing direction, since the source of information that allowed subject to directly determine the direction of the viewing vector has been removed. Experiment 1 examines this possibility.

# EXPERIMENT 1

## Methods

The eight paid subjects who participated in the experiment viewed a spatial layout made from a ground-plane reference and two slowly and irregularly tumbling wire-frame cubes marking positions on the reference and target positions on the plane. The techniques of data collection and viewing and display of the geometric projection were made identical to those used in previously described analytical and experimental studies (McGreevy and Ellis, 1986; Grunwald and Ellis, 1986)

A ground reference of irregularly spaced, parallel lines aligned with the reference direction was constructed with randomized spacing (fig. 1). To assure presentation of the correct lines of sight, the subject's eye was located at the center of projection. Two symmetrically placed viewpoint locations which were rotated clockwise and counterclockwise 22° with respect to a reference direction were used (left stations: -22°; right station: 22°). Both had a depression of 22° below the horizon. The target cubes were randomly placed at 72 equally spaced target azimuths. The subject showed his or her estimates of the target cube azimuth angle with respect to the reference direction by controlling a dial drawn electronically on the CRT with the method of adjustment.

## Results

Analysis of variance of the errors in target azimuth showed a statistically significant interaction between viewing station and true azimuth, (F = 2.413, df = 71,497, p < .001); hence, the azimuth error curves of left and right station appear to depend upon viewpoint.

Figure 2 shows the overall average error in the azimuth angle estimate for the left and for the right station plotted on circular graphs in which the direction of the error is shown as a directed arc. The across-subject means are good summaries of the data since the standard errors were only 1-4°. For both stations a systematic relationship between the azimuth error and the true azimuth angle is clearly recognized. Local minima in the errors, which are indicated by reversals in the directions of the error arcs, are not exactly where an actual grid-crossing axis would be, but are somewhat shifted toward a position orthogonal to the viewing axis. The largest direction errors are near ±45° and ±135° azimuth, and the error patterns for the symmetrically placed view stations are themselves approximate mirror images.

## Discussion

The symmetrical pattern of mean error clearly shows a dependency on view direction and demonstrates a breakdown of position constancy in the error pattern, thus confirming the initial hypothesis that removal of the crossing axis should break down the position constancy. This breakdown is particularly evident near ±90 target azimuth since these are generally not minimums as they were for previous experiments with gridded ground references (McGreevy and Ellis, 1986; Grunwald and Ellis, 1986). Thus, it is likely that the subjects are at least partially responding to the actual projected geometric properties of the scene which are seen from the separate viewpoints.

The breakdown of position constancy would be consistent with an alternative hypothesis which arises from previous analyses of errors in estimation of depicted directions in pictures (Ellis, Smith, and McGreevy, 1987; Gogel and Da Silva, 1987), and raises the classical question of the extent to which perception of an object's true geometric properties can be made to depend upon its projected retinal image (Thouless, 1931; Beck and Gibson, 1955; Gilensky, 1955; Gogel and Da Silva, 1987). According to this hypothesis, errors in judged direction in pictures are modeled as functions of the interrelations of actual lines of sight to viewed objects. For viewing situations in which pictures are viewed from the geometric center of projection, this analysis may be restricted to hypothesizing that the error in estimated target azimuth e is proportional to the difference between the depicted and projected azimuth angles y and y', respectively, i.e., $e = k(y' - y)$. Here the depicted angle y is measured with respect to the reference direction, clockwise positive, and the projected angle on the retina y' is measured with respect to the corresponding projection of the reference direction, clockwise positive. Positive errors correspond to clockwise errors. This formulation makes clear that not only should viewing direction affect the pattern of direction estimation, but also that symmetrically placed viewpoints should produce symmetrical patterns of direction errors.

The actual error data departs in significant ways from that expected based on this hypothesis. For example, the hypothesis implies that all direction errors for a view from the left station should be clockwise (fig. 3). The actual error data corresponding to this condition are both clockwise and counterclockwise, as shown by the circular plots of the error data. These error data could be modeled, as previously suggested, by introducing a 22° shift which produces an appropriate vertical shift in the theoretical function (McGreevy and Ellis, 1986; McGreevy, Ratzlaff, and Ellis, 1985). But this shift would be equivalent to asserting that the subject is responding to a potential projection rather than the one he or she actually sees. Since the data show evidence of viewpoint dependence and symmetry, the use of a theoretical function that suggests position constancy in the error data seems inappropriate. Accordingly, alternative theoretical explanations may be sought.

## Binocular Conflict

One possible influence on the direction judgements that the subjects were requested to make is the binocular stimulus which they viewed. This stimulus was essentially the picture surface which provided fixed accommodative and vergence demands as well as disparity and motion parallax cues to its physical distance. These cues tell the viewer that all objects are at an approximately equal egocentric distance, i.e., on the picture surface. Thus, if exocentric direction were to be based solely on egocentric ranges estimated from the binocular information, all targets would be at the same distance. In the reference system used, all targets would appear at azimuth positions perpendicular to the view direction; e.g., for a left view station they would appear either at 112° or 68°.

This binocular information is at odds with the monocular information that is drawn on the display, e.g., the size changes of the cubes as its depicted distance changes. The viewer is in a sense being presented with two simultaneous but conflicting stimuli, one binocular and the other monocular. One may suppose that the resulting perception is a combination of the two. Conflicts of this type have been studied in classical experiments (Beck and Gibson, 1955; Gogel, 1977) in which monocular and binocular stimuli are superimposed and viewed. Significantly, the finding has been that for some simple stimuli, the binocular depth sensation spreads to determine the apparent position of the monocularly viewed component of the visual field. Accordingly, it is

reasonable to suspect a similar process acting in the present experiment in which the binocular information in the picture surface causes the apparent positions of all targets to be attracted to a plane normal to the view direction. This process provides a hypothetical mechanism of the equidistance tendency observed in previous experiments. Its effects could be expected to be dominating were it not for the opposing influence of the remaining monocular depth cues provided by familiar shapes in the image.

## Familiar Shape

Assumptions regarding the physical properties of objects in pictures are necessary for picture perception because of the inherent ambiguity of the pictorial information. Though the images used for Experiment 1 are relatively impoverished in this respect, the viewer may introduce useful assumptions such as that the reference lines dropped from the cubes markers are parallel and equal and are themselves perpendicular to the ground reference. Other important assumptions would be that the marker cubes remain equal in depicted size and that the lines in the ground reference are all parallel and coplanar.

These assumptions allow the clarification of the ambiguities inherent in the picture and can account for residual viewpoint-independent aspects of the errors. For example, despite the absence of a crossing axis, the pattern of mean direction error reported reverses direction in a manner similar to that found in earlier experiments with gridded ground references. This judgement bias has been described as an "equidistance" since the errors indicated the perceived space is collapsed toward the crossing axis, compressing the space in a picture. The clear observation of this bias without a crossing axis shows that the crossing axis itself cannot be its cause.

Inspection of the circular plots of the direction error in figure 3 shows that zero crossings of the direction error are not as closely associated with the ±90° target positions in the present experiment as they were in similar experiments using a complete grid. In fact, there is substantial error at these positions. For the most part the actual zero crossings are along axes rotated towards positions orthogonal to the direction of view and hence parallel to the surface of the picture. That they are not completely rotated orthogonal to the view vector is probably due to distance cues based on the changing sizes of the cubes and reference lines which both provide relative distance information.

In fact, it is probably correct to argue that shape assumptions are the principal basis for the construction of a perceived space from the line-of-sight information provided by a picture. The properties of this inferred virtual space are opposed, however, by the properties of the physical space of the picture surface which, as mentioned earlier, provide a mechanism to produce the pattern of direction errors that have been recorded. A simple test of this hypothetical mechanism would be to repeat the previous experiment in a real scene, a situation where there is no binocular conflict. Experiment 2 investigates this possibility.

# EXPERIMENT 2

## Methods

Eight paid subjects viewed physical objects with the viewing geometry used in Experiment 1. The marker cubes were physically reproduced with PVC pipe and positioned in a parking lot adjacent to the Life Science Building at the Ames Research Center. The details of data collection and stimulus presentation are contained in a San Jose State University thesis (Smith, 1986). Conditions in Experiment 1 were generally duplicated, although electronically produced apertures and dials were replaced by actual objects with similar functions. A microcomputer randomized the sequence of conditions for each subject and timed and collected the responses.

The subjects viewed the stimulus scenes binocularly from about 61 cm behind and centered in the viewing windows. At the 28-m viewing distance the reference cube subtended an average 5.2°. The cubes markers provided a significant stereoscopic stimulus since the binocular disparity of the target varied between 6.6 to 9.8 ft around the reference cue. This maximum disparity difference of 3.2 ft is about 50 times the stereo threshold, but within typical values for fusion area for the retinal excentricities used. Subjects were required to make azimuth judgments of 24 equally spaced, randomly presented target positions. Two viewing directions (±22° left and right viewing stations, respectively) and two square window apertures (30° and 60° FOV) were used. The dependent variable again was the error in judging target azimuth direction.

The distance between the two observation stations was 21 m. Rather than have subjects walk this distance as often as a completely random schedule would dictate, each subject stayed at one direction of viewing for at least 16 trials (one block). For each direction of viewing, the factorial combination of 24 target cube directions, two window sizes, and two repetitions were randomly assigned to six blocks of 16 trials. Each subject was presented with 12 blocks of trials (six at each direction of viewing). The total of 192 trials required about 3 hr to complete.

## Results

The azimuth error data were analyzed by variance with repeated measures on target azimuth, window aperture, and viewing direction. Variation in the amount of background information by changing window size did not significantly affect judgments of azimuth error and did not interact with any other factor. As in Experiment 1, the two-way interaction between azimuth of the target cube and view direction was statistically significant ($F(23,138) = 3.861$, $p < .001$).

The nature of the statistical interaction that was observed between viewpoint and target azimuth is clarified by circular plots in figure 4. This figure illustrates the underlying symmetry in the error data, which is similar to that in Experiment 1. It also shows the absence of the "equidistance tendency" and generally smaller size of the errors.

## Discussion

The azimuth error observed in Experiment 2 does not exhibit the "equidistance tendency." Thus the results confirm the supposition that the binocular conflict or other cues to the picture surface such as motion parallax could be the cause of the bias. In Experiment 1 the azimuth errors for displays viewed from the correct geometric eye point were generally away from the reference axes and towards the crossing axis. This equidistance tendency has been called a "telephoto bias" since it resembles the pattern of error that would be induced if the view of the spatial configuration were distorted by a telephoto lens. In fact, it was not a true "telephoto bias" and equidistance tendency is a better description because the reported spatial compression was not aligned with the actual view direction, but with the axes, or implicit axes, in the scene itself. In contrast to the relatively large bias in Experiment 1, the errors in Experiment 2 are smaller and away from the crossing axes rather than towards them. The residual error pattern, however, does continue to exhibit a symmetrical dependence on view positions, supporting the conclusion from Experiment 1 that the error pattern does not exhibit position constancy. The new error pattern in Experiment 2 needs an explanation.

The bias pattern is not similar to what would be expected if it were due to the difference between the size of the projected and depicted azimuth angles. If the difference between depicted and projected angle were the cause of the observed error, the errors would be expected to resemble the traces in figure 3. As in Experiment 1, the results do not closely resemble these traces, so new alternatives need to be considered to explain both the smaller average size of the error and the particular pattern itself.

Since correct three-dimensional interpretation of the array of lines of sight to the objects in view depends upon both a correct internal model and a correct estimate of viewing direction, errors in either of these assumptions can be a source of systematic bias . Systematic errors in the internal model would result in apparent loss of perceptual rigidity when the object was rotated or translated. These kinds of distortions are not expected and were not reported as the cubes tumbled in the wind during Experiment 2. Accordingly, the biases found in this experiment might be attributed to incorrect estimation of the viewing direction. A classical error of this kind is called "slant overestimation" (Sedgwick, 1986) and corresponds to overestimation of the amount of depression of the viewing vector.

Figure 5 shows a family of theoretical azimuth error curves for different overestimates of the viewing vector depression together with the data from Experiment 2. These curves are constructed on the assumption that the viewer makes an error in the interpretation of the projected target angle, in a sense, by looking up its 3D characteristics in the wrong table. For example, the trace labeled "elevation = -40" shows the expected azimuth errors from a subject who, when looking a scene from a left viewing station (azimuth = 22.5°) with a -22.5° elevation angle, assumes that the actual elevation is -40°, and looks up the 3D interpretation of the projected angles that he or she does see in the wrong table, i.e., the one for a -40° elevation.  Interestingly, the hypothesis that azimuth error could be influenced by the difference between depicted target angle and its projection, which was described in the discussion of  Experiment 1, is really a special case of this kind of slant overestimation. The hypothesis discussed in Experiment 1 is equivalent to asserting that the overestimation is equal to the complement of the actual depression angle.

Figure 5  also shows the azimuth error data from Experiment 2 combined for both view stations by reflecting the data from the right view station so as to allow averaging with that of the

left station. The combined data are then replotted in cartesian form for comparison with the theoretical curves. The experimental data exhibit several features inconsistent with a slant overestimation. In particular, the errors are smaller, not markedly sinusoidal, and not biased in the correct directions. The elevation overestimation hypothesis predicts, for example, that from the left viewing station, errors for depicted angle between $0°$ and $180°$ should be clockwise whereas the data show a predominant counterclockwise bias for these conditions. In fact, the data may suggest an elevation underestimation. Clearly, further experiments in which errors in exocentrically judged azimuth and estimates of viewing direction elevation and azimuth are both collected are needed to evaluate the role of viewing direction misjudgement as an explanation for the pattern of azimuth error.

## Summary

1. Errors in exocentric judgements of the azimuth of a target generated on an electronic perspective display are not viewpoint-independent, but are influenced by the specific geometry of their perspective projection.

2. Elimination of binocular conflict by replacing electronic displays with actual scenes eliminates a previously reported "equidistance tendency" in azimuth error, but the viewpoint dependence remains.

3. The pattern of exocentrically judged azimuth error in real scenes viewed with a viewing direction depressed $22°$ and rotated $\pm22°$ with respect to a reference direction could not be explained by overestimation of the depression angle, i.e., a slant overestimation.

# REFERENCES

Attneave, F., and Frost, R. (1969) The determination of perceived tridimensional orientation by minimum criteria. *Percept. and Psychophys., 6,* 391-396.

Beck, Jacob and Gibson, James J. (1955) The relation of apparent shape to apparent slant in the perception of objects. *J. Exp.Psychol.,* 50, 125-133.

Borresen, C. Robert and Lichte, William H. (1962) Shape constancy: dependence upon stimulus familiarity. *J. Exp.Psychol., 63,*92-97.

Ellis, Stephen R., Smith, Stephen, and McGreevy, Michael W. (1987) Distortions of perceived visual directions out of pictures. *Percept. Psychophys.,* 42, 535-544.

Gilinsky, A.S. (1955) The effect of attitude upon perception of size. *Amer. J. Psychol., 68,* 173-192.

Gogel, Walter C. (1977) The metric of visual space. in *Stability and constancy in visual perception,* W. Epstein ed. New York, Wiley.

Gogel, Walter C. and Da Silva, Jose Aparecido (1987) Familiar size and the theory of off-sized perceptions. *Percept. Psychophys., 41,* 318-428.

Goldstein, E. Bruce (1987) Spatial layout, orientation relative to the observer, and perceived projection in pictures viewed at an angle. *J. Exp.Psychol., 13,* 256-266.

Gregory, Richard L. (1966) *Eye and brain: the psychology of seeing.* New York: World University Library

Grunwald, Arthur, and Ellis, Stephen R. (1986) Spatial orientation by familiarity cues. in *Training, human decision making and control.* J. Patrick and K.D. Duncan, eds. North-Holland, Elsevier, 257-279.

Ittelson, W.H. (1951) Size as a cue to distance. *Amer. J. Psychol., 64,* 54-67.

Ittelson, W.H. (1952) *The Ames demonstrations in perception,* Princeton: Princeton University Press.

Hochberg, Julian. (1987) Machines should not see as people do, but must know how people see. *Computer Vision, Graphics, and Image Processing* ( in press).

Howard, Ian (1982) *Human Visual Orientation,* New York: Wiley

Flavell, J.H. (1963) *The developmental psychology of Jean Piaget.* Princeton: Van Norstand.

Metzger, Wolfgang (1953) *Gesetze des Sehens,* Frankfort am Main: Waldeemar Kramer

McGreevy, Michael W., Ratzlaff, Cordell, and Ellis, Stephen R. (1985) Virtual space and two-dimensional effects in perspective displays. *Proc. 21st Ann. Conf. on Manual Contr.,* June, 1985.

McGreevy, Michael W. and Ellis, Stephen R. (1986) The effect of perspective geometry on judged direction in spatial information instruments. *Human Factors,* 28, 439-456.

Sedgwick, H. A. (1986) Space perception in *Handbook of Perception and Human Performance vol 1.* K. R. Boff, L. Kaufman, J.P. Thomas eds, New York: Wiley, 21:19 - 21:23

Piaget, J. (1932) *The origins of intelligence in children,* London: Keagan Paul.

Pirenne, M. (1970) *Optics, painting, and photography,* Cambridge: Cambridge Univ. Press

Smith, Stephen (1986) Viewing in three-dimensional visual space: the two-dimensional effect and direction judgement errors. Unpublished MA thesis, San Jose State Univ., San Jose, CA.

Thouless, Robert H. (1931) Phenomenal regression to the real object. *British J. Psychol., 21,* 338-59.

Wallach, Hans. (1985) Perceiving a stable environment. *Sci. Amer., 252,* 118-124.
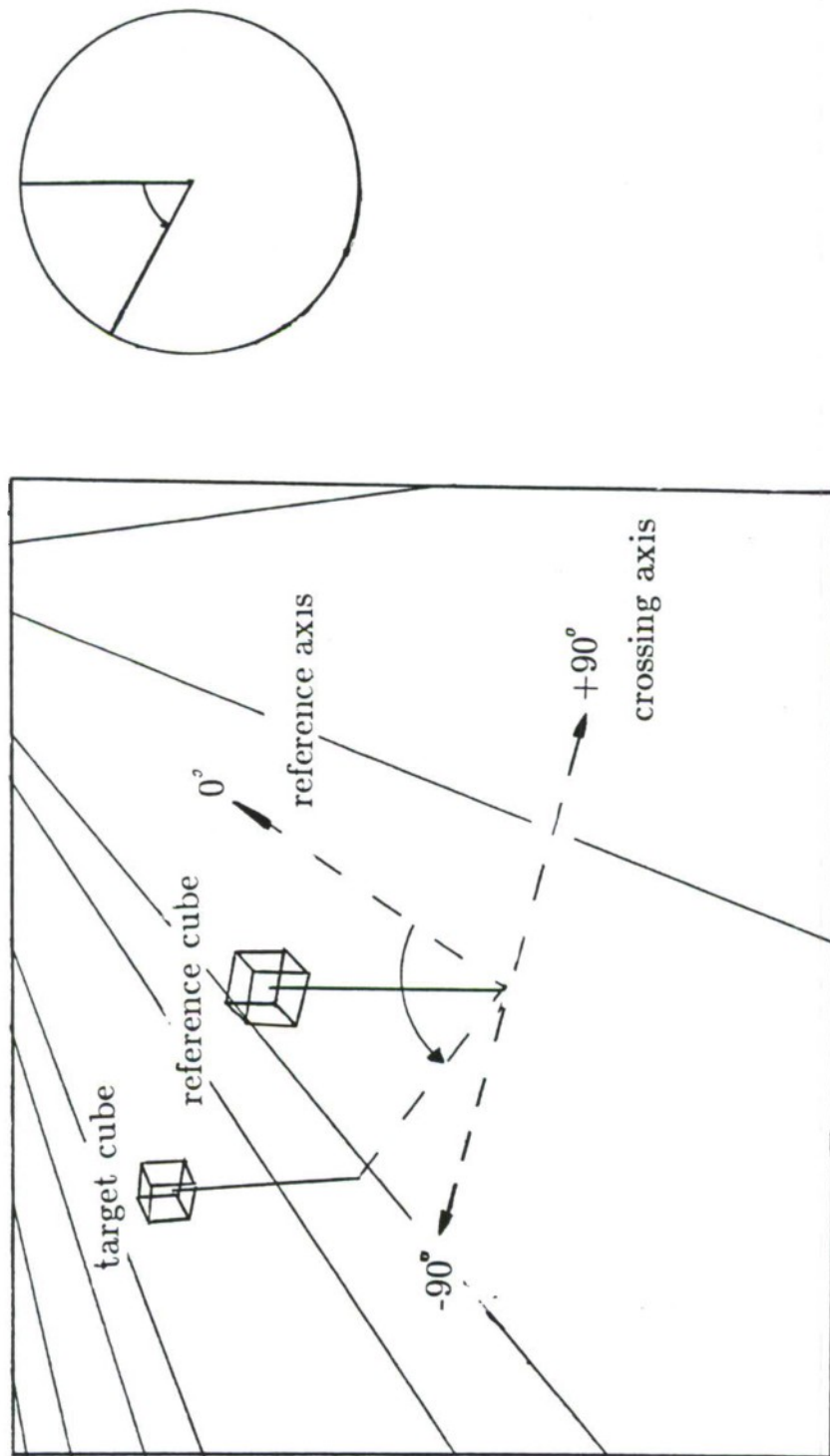
Figure 1. Schematic of the direction judgement task. The subject adjusted the angle shown on the dial at the right until it appeared equal to the azimuth angle of the target cube. Dashed line, labels, and arrows did not appear on the display. The ground reference in previous experiments was a full grid.
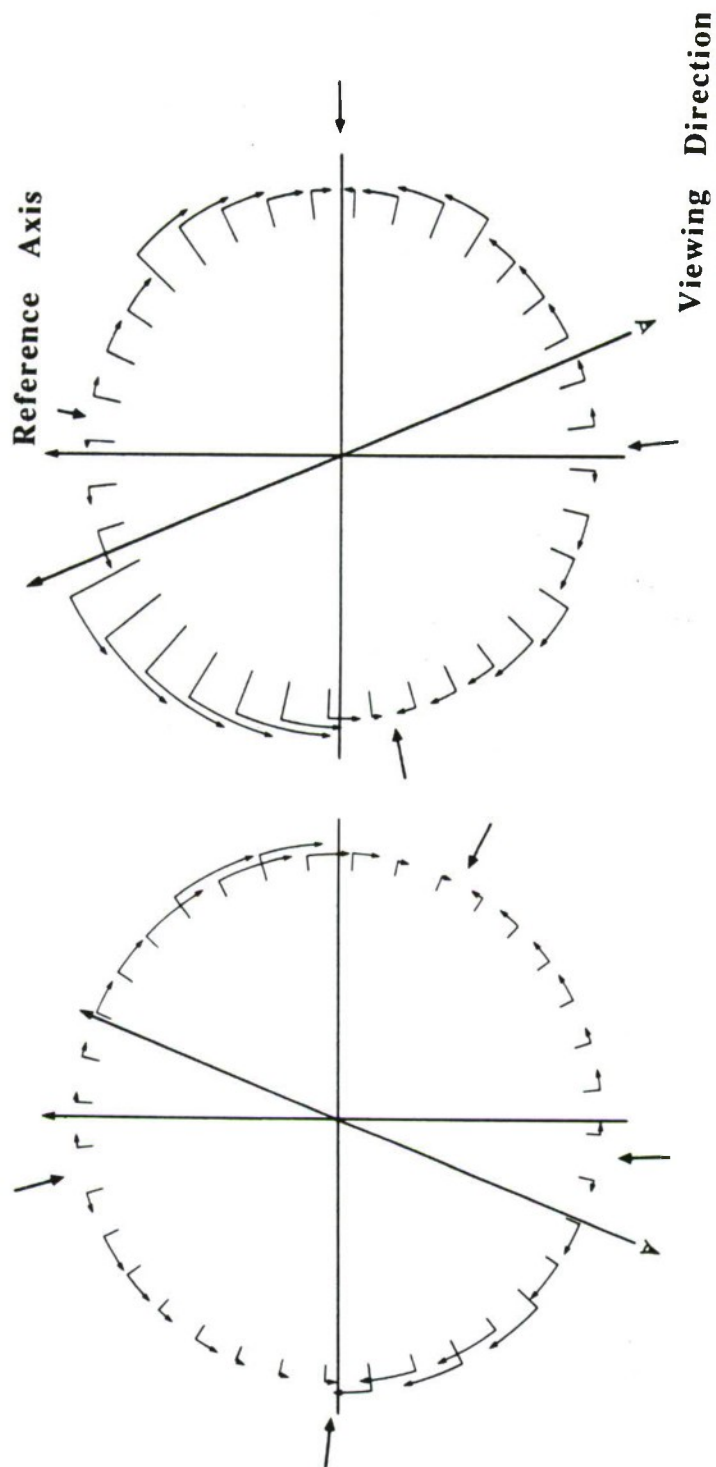
Figure 2. Circular plot of mean azimuth error for direction judgement experiment using a display with a ground reference of parallel lines. The length of the arc corresponds to the mean error in estimating target azimuth. Arrows show target azimuths, where azimuth errors were at local minimum.
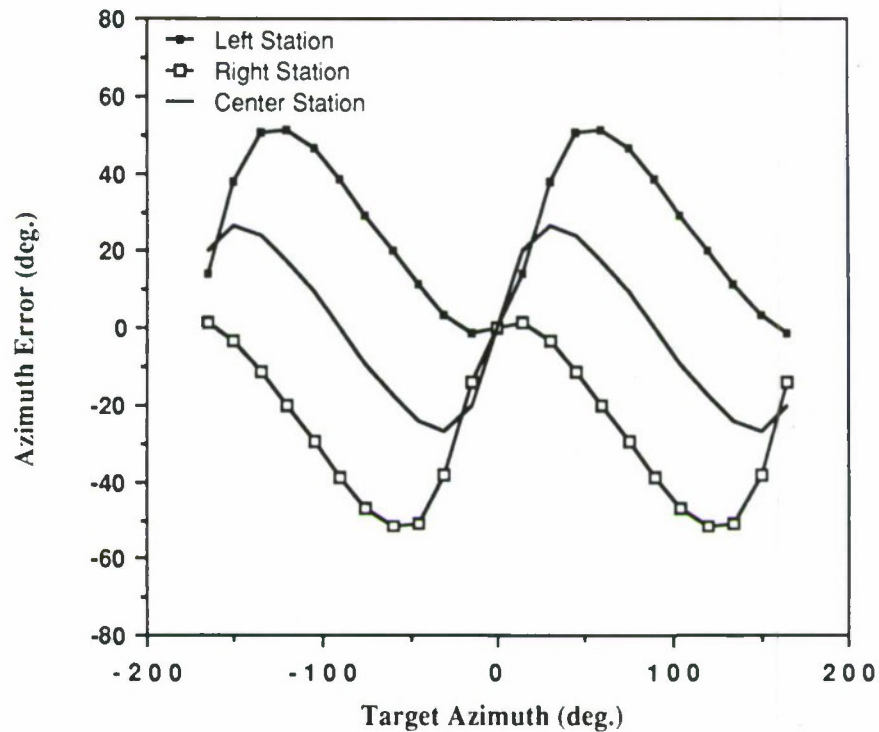
Figure 3. Predicted azimuth errors. If the subjects' direction errors were entirely determined by the difference between the true depicted value of a target's azimuth angle and its projection, errors like those shown in this figure would be expected. The three traces show the expected error pattern if the depicted targets were viewed from a left (22.5°), right (-22.5°), or center (0°) viewing azimuth.
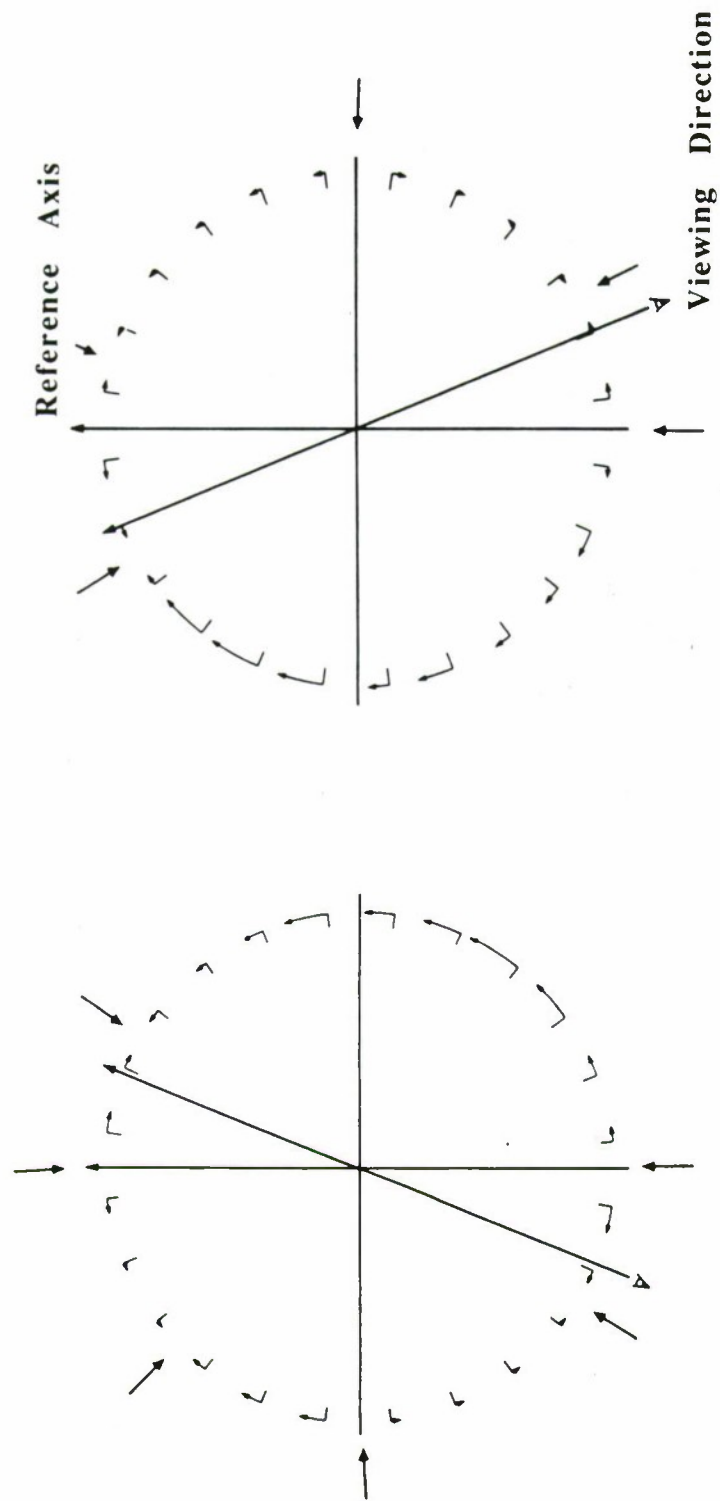
Figure 4. Circular plot of mean azimuth error for direction judgment experiment using an actual scene. Arrows show target azimuths where azimuth errors were at local minimum.
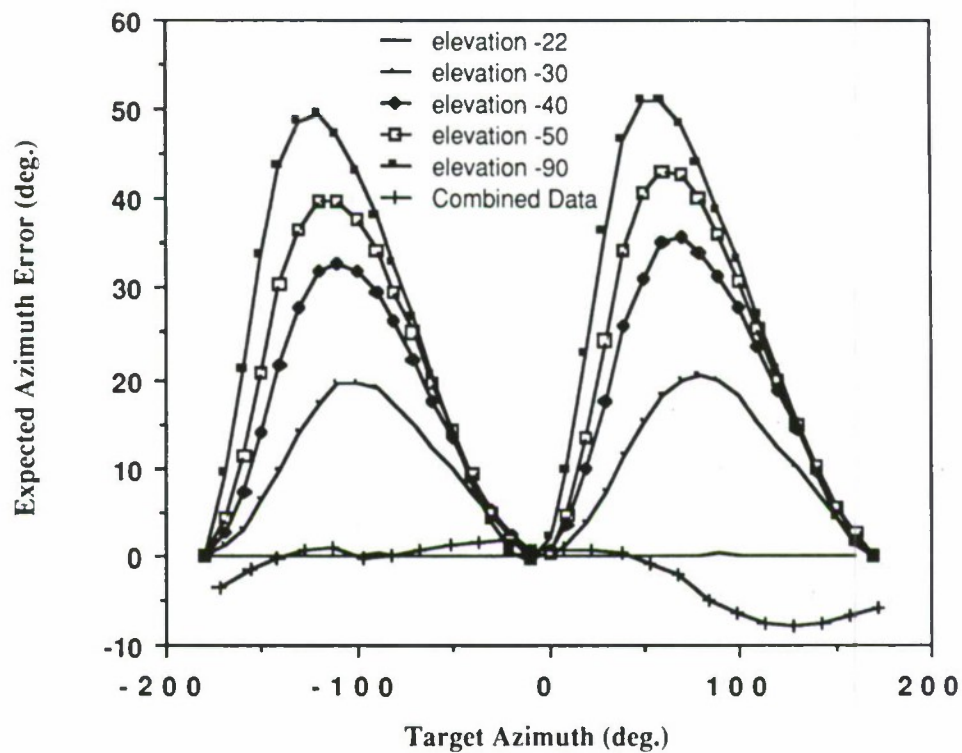
Figure 5. Plot of expected azimuth error if a subject misjudged the depression angle of the viewing direction. Errors are calculated for a left viewing station (azimuth = 22.5°) with a depression angle of -22.5°, assuming that the subject misjudged the depression to be the parameter of each of the curves. Error data from Experiment 2 are also plotted for comparison.

# HOW TO REINFORCE PERCEPTION OF DEPTH IN SINGLE TWO-DIMENSIONAL PICTURES*

S. Nagata
NHK Science and Technical Research Laboratories
Tokyo, Japan

## ABSTRACT

The physical conditions of the display of single two-dimensional pictures, which produce images realistically, were studied by using the characteristics of the intake of the information for visual depth perception. "Depth sensitivity," which is defined as the ratio of viewing distance to depth discrimination threshold, has been introduced in order to evaluate the availability of various cues for depth perception: binocular parallax, motion parallax, accommodation, convergence, size, texture, brightness, and air-perspective contrast. The effects of binocular parallax in different conditions, the depth sensitivity of which is greatest at a distance of up to about 10 m, were studied with the new versatile stereoscopic display. From these results, four conditions to reinforce the perception of depth in single pictures were proposed, and these conditions are met by the old viewing devices and the new high-definition and wide television displays.

## I. INTRODUCTION

The sensation of reality in a picture occurs because of visual depth perception. Therefore, in order to display pictures as if the observer were looking at real objects in three-dimensional space, the physical conditions of the pictures must be matched to the characteristics of the process involved in the intake of information relative to depth perception. The objectives of this paper are to report the results of an investigation on the availability of many cues for visual depth perception, using a common evaluating scale, and to propose ways to reinforce the perception of depth in single two-dimensional pictures.

It is well known that a pair of pictures taken from two laterally separated positions creates the effect of stereoscopic depth perception with binocular cues, such as binocular parallax and convergence cues of the eyeball shown in figure 1 and table 1. However, there are other monocular cues shown in figure 1, such as the accommodation cue of a crystalline lens, motion parallax on moving vision, and pictorial cues. The pictorial cues include transversal size, longitudinal size, texture density and shape, intersection, position of horizon, brightness and shade, air-perspective, and color effect. The study of the comparison of the effectiveness of each of the cues and the study of the interaction between different cues are necessary.

---

The availability of cues for visual depth perception has been investigated. Künapas (ref. 1) studied the subjective absolute distance by the method of magnitude estimation as a function of viewing distance up to 4 m and five viewing conditions, where the cues (retinal size, binocular parallax, accommodation, and brightness) were fully provided or partially reduced.

He found that accommodation did not permit any accurate perception of distance, and that retinal image size was one of the most important cues in the judgement of absolute distance from the observer. He also pointed out the similarity of his result and the result of Holway and Boring (ref. 2) that the apparent size at a fixed viewing distance varies with the viewing condition. However, Künapas did not study motion parallax and relative depth perception.

When we view a picture which contains many objects, the space perception in the picture depends on the results of the relative depth perception among the objects.

Stubenrauch and Leith (ref. 3), using holograms, found the interposition cue to dominate over most combinations of other cues (binocular parallax, motion parallax, and retinal size) for perception of normal relief or reversed relief. However, these effectiveness estimations were not measured at large viewing distances.

Furthermore, since the cues on the retina, such as parallax, size, brightness, etc., have different physical attributes, the threshold value of each cue change for depth perception cannot be directly compared with each other.

The author proposes a common scale for evaluating the availability of depth cue, which is defined as the ratio $D/\Delta D$ of the viewing distance D to the detection threshold $\Delta D$ of depth difference (depth threshold). We call this ratio scale "depth sensitivity" (refs. 4,5) of vision.

In this way, the effectiveness of various cues can be quantitatively compared with each other as a function of viewing distance.

## II. METHODOLOGY

### Hypothesis

First, the relationship between depth sensitivity and the detection of quantitative cue change for depth perception of the object's image on the retina was considered from the viewpoint of the hypothesis that the change of cues is transformed into perception depth information while at the same time conserving the information concerning the character of the object on the base of the character as shown in table 1.

For example, when a value R(D) of the cue of binocular viewing direction is inversely proportional to the viewing distance D and is proportional to the constant A (where A is the distance between two pupils), the detection threshold $\Delta R$ of the change of the value of a cue such as the binocular parallax is obtained from the depth threshold $\Delta D$ as follows:

$$\Delta R = R(D) - R(D + \Delta D) = A/D - A/(D + \Delta D) \tag{1}$$

Then the depth sensitivity is deduced as

$$D/\Delta D = A/(\Delta R \cdot D) - 1 \qquad (2)$$

Second, by dilating on Fechner's Law (ref. 6), it was proposed that depth sensation is based on the sum of the small depth sensation unit $dS=K$ corresponding to the depth thresholds. The depth sensation $S(D)$ is obtained by

$$S(D) = \int_{D_0}^{D} \left(\frac{dS}{dD}\right)dD = \int_{D_0}^{D} (K/\Delta D)dD \qquad (3)$$

where K is a transformation constant.


## Psychophysical Experiment

The depth sensitivities of the cues of binocular parallax, motion parallax, and accommodation were obtained from the depth thresholds in psychophysical experiments. The characteristics of the cue-change threshold $\Delta R$ is induced from the depth threshold $\Delta D$ measured under the limited condition, and the depth sensitivities of these cues were calculated. Furthermore, the depth sensitivities of other cues were also calculated by estimating the detection threshold of cue change.


## III. EXPERIMENTS I


For measuring the depth threshold, an observer, by using a remote-wire system, moved one of the two black rods (20 arc/min in width, 1 cd/m$^2$) as illustrated in figure 2, so the difference of depth can be noticed through a slit (40 arc/min in height, 19 arc/deg in width).

Two males (SN 33 yr of age, left V.A. 1.2 corrected, right V.A. 1.2; KI 23, 1.5, 1.5) and one female (NW 23, 1.2, 0.6) having normal streoscopic vision served as subjects in these experiments.

The depth thresholds on the binocular parallax, the motion parallax, and the cue from accommodation were measured as a function of viewing conditions.

The viewing conditions for controlling the depth cues were obtained by combining binocular observation or monocular observation and static observation or lateral moving observation, and observation with natural pupils or artificial pupils (1 mm diameter).

In moving observations, the observer moves the upper body rhythmically to the right and left at different distances and velocities which were measured in real time by an electronic scale wired to the head.

The viewing distance to the fixed rod was 1, 2, 5, and 18 m, respectively. The brightness, the retinal size, and the interval distance of two stimuli were not changed as a function of observation distance.

The measurements were taken for eight trials a day for three days for each person under each condition.

## IV. RESULTS I

### Binocular Parallax, Motion Parallax, and Accommodation Cues

The depth thresholds with static binocular vision through natural pupils were obtained as shown in table 2, and the symbol (o) in figure 3 indicates the depth sensitivities as a function of distance obtained from the depth threshold of the typical subject (SN).

The cue-change threshold $\Delta R$ on binocular parallax shown in figure 4 was calculated from the depth threshold $\Delta D$ with binocular vision from equation (1). It was considered that the binocular threshold neither changed as a function of viewing distance, i.e., convergence angle, nor as a function of the size of the pupils.

This was in agreement with the other two observers' results and with those reported by Ogle (ref. 7), Zoth (ref. 8), and Nishi (ref. 9). But Amigo (ref. 10), and Lit and Finn (ref. 11) reported that the threshold slightly increases as the distance decreases to less than 1 m because of the instability of the oculomotor.

The depth sensitivities of this cue shown by the solid line in figure 3 are calculated from equation (2), where a constant value is substituted for $\Delta R$. The maximum distance $D_{max}$ for which the sensitivity falls to zero is $A/\Delta R$.

The depth sensation $S_D$ on binocular parallax is deduced from equation (4) and may be saturated at about 10 m (ref. 12).

$$S_D = \int_{D_0}^{D} \left(\frac{K}{\Delta D}\right) dD = K \int_{D_0}^{D} \left(\frac{A}{\Delta\theta \cdot D^2} - \frac{1}{D}\right) dD$$

$$\text{(4)}$$

$$= K\left[\frac{A}{\Delta\theta}\left(\frac{1}{D_0} - \frac{1}{D}\right) + \log_n \frac{D_0}{D}\right]$$

The depth thresholds for motion parallax with moving monocular vision with a natural pupil at the speed at which the subject could detect the depth are shown by the symbol (■) in figure 3. The depth threshold at a viewing distance of 3 m was measured for different conditions, and it was dependent on the velocity $\omega_D$, but not on the distance d of movement as shown in

figure 5(A). The optimum velocity $\omega a$ was 6-8° of arc/sec, and at velocities lower than the optimum velocity, the threshold velocity of motion parallax is constant.

Graham (ref. 13) and Zeger (ref. 14) reported on the increase in the threshold as the velocity increases from about 6° to 20° of arc/sec. But in our results shown in figure 5(B), the velocity threshold of motion parallax $\Delta\omega$ is constant at velocities lower than the optimum. This constancy is deduced from the detection model where the minimum parallax is sampled at a constant interval time.

The depth sensitivity of motion parallax calculated from equation (2) is represented by the solid line in figure 3. The sensitivity is $\omega a/\Delta\omega$ and is constant up to the distance at which the optimum velocity of the body movement is obtained, and when the distance is exceeded the sensitivity decreases. The descending curve is obtained by substituting the maximum velocity $V_{max}$ of body movement for A, and $\Delta\omega$ for $\Delta R$ in equation (2).

This motion parallax is produced not only by the absolute motion of the observer, but also by the relative motion of the objects, and in the case of moving vision on some riding machine with a speed higher than the motion of the body, the sensitivities of motion parallax at large distances are maintained at the same level as that for short distances and are higher than that for binocular parallax.

The depth thresholds for the blurring cue of accommodation with static monocular vision through a natural pupil are represented by the symbol ($\Delta$) in figure 3, and the depth threshold with vision through the artificial pupil was nearly equal to or slightly greater than the viewing distance.

The depth sensitivity of the natural accommodation cue was also calculated by substituting the pupil diameter during observation for A in equation (2) and by substituting the blurring threshold resulting from equation (1) (similar to the reciprocal of his visual acuity) for $\Delta R$ in equation (2).

## Other Cues

The depth sensitivities relative to binocular parallax, motion parallax, and accommodation cues obtained from equation (2) are satisfied by the data resulting from the experiment. Therefore, we applied the same method of analysis in obtaining this sensitivity data to the sensitivity data relative to the other cues: convergence, size, slanted shape, texture density, brightness, and air-perspective contrast.

In figure 2, when two objects positioned at a large visual angle are observed in binocular vision, convergence of the line of sight of two eyes results in depth perception. However, the detection threshold of convergence change is larger than the detection threshold of binocular parallax, and the depth sensitivity of convergence was obtained from equation (2) and is represented by dashed line in figure 3.

The depth sensitivity to the cue of the object transversal size shown in figure 3 was calculated from equation (2), where size-S is substituted for A and the ratio of the size change detection threshold $\Delta\theta(\equiv\Delta R)$ to size $\theta = S/D$ in visual angle is constant as reported by Ogle (ref. 15).

This sensitivity agrees with the depth threshold under monocular observation of two square targets (1.8 m$^2$) measured by Teichner (ref. 16). The maximum distance $D_m$ is determined by the absolute detection threshold of size perception.

The depth sensitivities on the shape of a rectangular object whose upper part inclines at larger distances is represented by

$$\frac{D}{\Delta D} = \frac{S}{\Delta\theta \cdot D} - 1 = \frac{D}{L \cdot \sin\alpha_t} \tag{5}$$

where S is the horizontal length of object, $\Delta\theta$ is the size-cue threshold, L is the height of object, and $\alpha_t$ is the slant threshold.

Freeman (ref. 17) measured the slant threshold of 14 different rectangles without texture by monocular vision. The depth sensitivities calculated from these data varied with height. The optimal depth sensitivity was 78 when D = 135 cm and L = 8 cm. This sensitivity is larger than the data of Teichner, resulting in the difference between the shape cue of one object and the size cue of two separate objects.

In viewing a textured pattern, there are different sizes or density of texture: one is the transversal size or density in a plane rectangular to the depth direction as mentioned above, and the other is the longitudinal size or density along the depth-directional line.

The depth sensitivity on the latter was calculated from equation (6) and is shown in figure 3:

$$\frac{D}{\Delta D} = 2\left(\frac{S \cdot H}{\Delta\theta \cdot D^2} - 1\right) = 2\left(\frac{\theta}{\Delta\theta}\right) \tag{6}$$

where S is the object's longitudinal size on the depth direction, H is the distance between the visual line and the object plane, and the ratio of the longitudinal size $\theta$ in visual angle to the size cue threshold $\Delta\theta$ is the same as the ratio of the transversal size cue threshold. This sensitivity is twice as large as that for the transversal size.

The depth sensitivity on the brightness cue shown in figure 3 is deduced from equation (7):

$$\frac{D}{\Delta D} = 2\frac{L \cdot r}{\Delta I \cdot D^2} = 2\frac{I}{\Delta I} \tag{7}$$

where L is luminous intensity, D is the lighting distance, r is the refractory factor of the object, I is the luminance of objects, and $\Delta I$ is the cue-change threshold of luminance. This sensitivity is satisfied even at a very small stimulus level at which point Ricco-Piper's law is applied.

When the observer or viewing objects move in three-dimensional space, the projected retinal image changes in position, size, shape (ref. 18), density, and luminance, and the depth perception is effected by those changing velocity.

The depth sensitivity of the air-perspective contrast cue results from the contrast-diminishing function of equation (8), except for the case of blurring or color effect.

$$C = C_0 \exp\left(-\frac{D}{\sigma}\right) \tag{8}$$

where $C_0$ is the luminance contrast at very small distances and $\sigma$ is the length constant determined by the air-scattering coefficient.

The sensitivity on this cue illustrated in figure 3 is calculated from equation (9):

$$\frac{D}{\Delta D} = \frac{-C_0}{\Delta C} \cdot \frac{D}{\sigma} \exp\left(-\frac{D}{\sigma}\right) = -\frac{C}{\Delta C} \cdot \frac{D}{\sigma} \tag{9}$$

where $\Delta C$ represents the differences in threshold for the brightness contrast deduced from the variation of detection threshold relative to the sine-wave grating pattern given by Watanabe et al. (ref. 19).

## V. EXPERIMENTS II

Because of the above-mentioned result that the depth sensitivity of binocular parallax was very high in comparison with other cues, the effects of binocular parallax in other conditions and the interaction effect between binocular parallax and other monocular cues were studied. In Experiments I the change of binocular parallax and retinal size corresponding to moving objects in depth could not be controlled independently. To measure the effects of two coexistent cues, the new versatile stereoscopic display (ref. 20) of the standard TV system in conjunction with a special video processor (fast phase modulation) were used.

In this system, as shown in figure 6, the stereoscopic pictures have been produced with binocular parallax and convergence, controlled temporally and spatially with depth signals in a manner comparable to brightness control signals of video signals – all independent of pictorial cues; for example, size of pattern. The picture is also changed independent of the depth signal.

## VI. RESULTS II

The subjects viewed the square pattern in streoscopic vision, of which the size and binocular parallax was changed temporally and simultaneously by the pattern-size and depth-control sine-wave synchronous signals, with variable amplitude and polarity of depth direction, so that the conditions of equally felt depth sensations of motion could be measured. In figure 7, the horizontal axis represents the amplitude in arc-minutes peak-to-peak of oscillation of binocular parallax and the vertical axis represents the amplitude of oscillation of size. The smoothed curves indicate the conditions of those two cues for which equal depth sensation occurred at three levels; that is, depth threshold ($\Delta$) and two suprathreholds ($\bullet, \blacktriangle, \square$). The data show that the depth sensation from two coexistent cues, changing size and binocular parallax, is a combination of the individual effects of each cue, and when binocular parallax is zero, the changing size cue in

monocular vision is more effect than the changing size cue in binocular vision. In other experiments, it was found that the effect of binocular parallax decreased when the objects moved in depth or in the lateral direction.

## VII. DISCUSSION AND CONCLUSIONS

The following conclusion were derived from the comparison of the depth sensitivities of various cues and from the interactive effects of depth sensation from two different cues, size changing and binocular parallax.

1. The depth sensitivity relative to binocular parallax is maximum at a distance of up to about 10 m.

2. The depth sensitivity to motion parallax is effective, and this sensitivity on motion at the optimum velocity exceeds that of the binocular parallax at a distance greater than 10 m.

3. The cues from accommodation and convergence are effective for the relative depth perception only at a distance of less than 1 m, but are effective for the absolute depth perception at longer distances.

4. The pictorial cues are effective even at long distances, and the sharp edge of pictures, and clear texture, shade, and gloss of the surface on objects strengthen the sensation of depth.

5. The effects of these cues work together and combine spatially on the wide visual field.

From the investigation of these sensitivities, the following conditions to decrease the sensation of flatness of the display plane of single two-dimensional pictures and to reinforce the depth perception in the picture were found:

1. The effects of binocular parallax must be decreased.

2. The distance of convergence and accommodation must be close to the actual distance of the objects in the picture.

3. The frame of the display must be separated from the images peripherally or depth-wise to be defused.

4. There must be many monocular pictorial cues including the projection of three-dimensional moving objects.

Conditions 1, 2, and 3 are attained by viewing with monocular vision or by positioning the picture image at a distance of about 5 m; conditions 3 and 4, by making the visual angle of picture wide; and condition 4, by using a hi-definition and moving picture.

So, we can point out that the new high-definition and wide television displays (ref. 21) meet these conditions, and these displays produce more realistic picture images than the conventional television.

It is well known that one of the importance conditions for space perception is the size of the viewing field of the display, which gives self-motion perception to an observer, such as when one stands in a "Wander-Room" where wall and ceiling surrounding one rotates; nevertheless, one feels self-motion.

It was found that a visual wide-angle display over 30° induces the sensation of reality because of the integration of the depth cue effects (ref. 22).

The old viewing device called reflectorscope or vue d'optique (in Japanese, nozoki-karakuri, which means "peeking device"), shown in figure 8, in which pictures were viewed through a convex lens or a concave mirror, produces images of the picture realistically.

Concerning the reasons why this device produces reality, Valyus (ref. 24) and Schwartz (ref. 25) pointed out that because of the aberration of the lens or reflector, binocular parallax occurs and results in stereoscopic pictures, and also the difference between the illumination intensities of the binocular images, because of the difference of the diffusion of the screen, results in strereoscopic vision. If these explanations are correct, the disparity and the difference of illumination between the binocular images would increase with the distance from the median line of the picture, and then the depth sensation would depend on position.

However, according to the results of our observations, the depth sensation depends on the nature of objects in the picture, and the depth sensation in monocular vision is equal to or better than that in binocular vision.

Therefore, the actual reason why reality is produced on the old viewing device is that they fulfill conditions proposed in our results in the case of pictures without movement.

# REFERENCES

1. Künapas, T. M.: Distance Perception as a Function of Available Visual Cues, J. Exp. Psy., 77, 1968, pp. 523-529.

2. Holway, A. H.; and Boring, E. G.: Determinants of Apparent Visual Size with Distance Variant, Am. J. Psychology, 51, 1941, pp. 21-37.

3. Stubenrauch, C. F.; and Leith, E. N.: Use of Hologram in Depth-Cue Experiments, J. Opt. Soc. Am., 61, 1971, pp. 1268-1269.

4. Nagata, S: Depth Sensitivities of Various Cues for Visual Depth Perception, J. Insti. TV. Engrs. Japan, vol. 31, no. 8, 1977, pp. 649-655 (in Japanese).

5. Nagata, S.: Visual Depth Sensitivities of Various Cues for Depth Perception, NHK LABS. NOTE, vol. 266, 1981, pp. 1-11.

6. Fechner, G. T.: Elemente der Psychophysik.Leipzig: Breikopf and Hartel (1889), in J. P. Guilford, Psychometric Methods (McGraw-Hill, New York, 1936), pp. 20-42.

7. Ogle, K. N.: Note on Streoscopic Acuity and Observation Distance, J. Opt. Soc. Am., 48, 1958, pp. 794-798.

8. Zoth, in Frobes J., Lehrhuch der Experimentellen Psychologie, Freiburg; Herder (1923).

9. Nishi, I.: Limits of Depth Perception, Proceedings IV of Psychology Tokyo, Iwanami Pub., 1933, pp. 160-165.

10. Amigo, G.: Variation of Stereoscopic Acuity with Observation Distance, J. Opt. Soc. Am., 53, 1963, pp. 615-630.

11. Lit, A.; and Finn, J. P.: Variability of Depth-Discrimination Threshold as a Function of Observation Distance, J. Opt. Soc. Am., 66, 1976, pp. 740-742.

12. Nagata, S.: Depth Information on Vision and Depth Sensation, Proc. Image Engineering Conf. 6 (5-2), 1975, pp. 81-84. (In Japanese).

13. Graham, C. H.; Baker, K. E.; Hecht, M.; and Lloyed, V. V.: Factors Influencing Thresholds for Monocular Movement Parallax, Exp. Psychol., 38, pp. 205-233, in Vision & Visual Perception, edited by C. H. Graham, John Wiley & Sons, New York, 1948, pp. 513-516.

14. Zegers, R. T. J. S.: Monocular Movement Parallax Thresholds as Functions of Field Size, Field Position and Speed of Stimulus Movement, J. Psychology, 26, 1948, pp. 477-498.

15. Ogle, K. N.: In The Eye IV, edited by H. Davson (Academic Press, New York and London), 1962, pp. 233-234.

16. Teichner, Kobrick, and Wehrkamp, Am. J. Psychology, 68, 1955, pp. 193-208.

17. Freeman, R. B.: Absolute Threshold for Visual Slant, J. Exp. Psychol, 71-2, 1966, pp. 170-170.

18. Hochberg, J.; and Brooks, V.: In The Psychology of Vision Perception, edited by R. N. Harber (Holt Inc., New York, 1960), pp. 195-197.

19. Watanebe, A.; Mori, T.; Nagata, S.; and Hiwatashi, K.: Spatial Sine-Wave Responses of the Human Visual System, Vision Res., 8, 1968, pp. 1245-1263.

20. Nagata, S.: New Versatile Sterio Display (NS Type) System and Measurement of Binocular Depth Perception, Japanese Journal of Medical Electronics and Bionengineering, vol. 20, no. 3, 1982, pp. 16-3.

21. Fujio, T.; Picture Display Systems for Future High-Definition Television, Proceedings of the 3rd International Display Res. Conference, 1983, pp. 554-557.

22. Hatada, T.; Sakata, H.; and Kusaka, H.: Psychophysical Analysis of the "Sensation of Reality" Induced by a Visual Wide-Field Display, SMPTE Journal, 89, 1980, pp. 560-569.

23. Ninomiya, Y.: Parabolic Mirror Made by the Rotation Method: Its Fabrication and Defects, Appl. Opt. 18, no. 11, 1979, pp. 1835-1841.

24. Valyus, N. A.: Stereoscopy, (Focal Press, London, 1966), pp. 80-82.

25. Schwartz, A. H.: Stereoscopic Perception with Single Pictures, Optical Spectra, September 1971, pp. 25-27.

## TABLE 1. DEPTH PERCEPTION CUES AND BASES OF CHARACTER FOR TRANSFORMATION. O:OCULOR MOTOR CUE. R:RETINAL IMAGE CUE. B:BINOCULAR VISION. M:MONOCULAR VISION.

| Cue | Objective change in 3-D | Image change in retina | Base of character for transformation |
|---|---|---|---|
| Binocular parallax | R.B.  Relative distance | Position disparity | Unity |
| Binocular convergence | O.B.  Absolute distance | Position | Optimum |
| Accommodation (blurring) | O.M.  Absolute | Blurring | Optimum |
| Motion parallax | O.M.  Absolute R.M.  Relative | Position disparity Velocity | Unity |
| Transversal size | R.M.  Relative Absolute (familiar) | Size | Identity |
| Longitudinal size | R.M.  Relative | Size | Identity |
| Vertical position | R.M.  Absolute | | |
| Size, density | R.M.  Slant in depth | Size, density | Uniformity |
| Shape | R.M.  Slant | Shape | Simpleness uniformity |
| Motion | R.M.  Motion | Velocity flow | |
| Intersection | R.M.  Front and Back | Shape | Simpleness |
| Luminance | R.M.  Relative | Illumination | Uniformity |
| Shade | R.M.  Slant | Illumination | Uniformity |
| Air-perspective | R.M.  Relative | Contrast, blurring color | Identity |
| Color | R.B.M.  Aberration | Color disparity | Unity |

TABLE 2.  DEPTH THRESHOLDS ON BINOCULAR VISION AS A FUNCTION OF
VIEWING DISTANCE

| Viewing distance, m | 2 | 3 | 5 | 18 |
|---|---|---|---|---|
| Sub. SN | 0.5 | 1.9 | 5.7 | 5.1 (cm) |
| Sub. KI | 0.2 | 0.4 | 1.2 | 2.9 |
| Sub. NW | 0.8 | 2.0 | 6.1 | 2.9 |
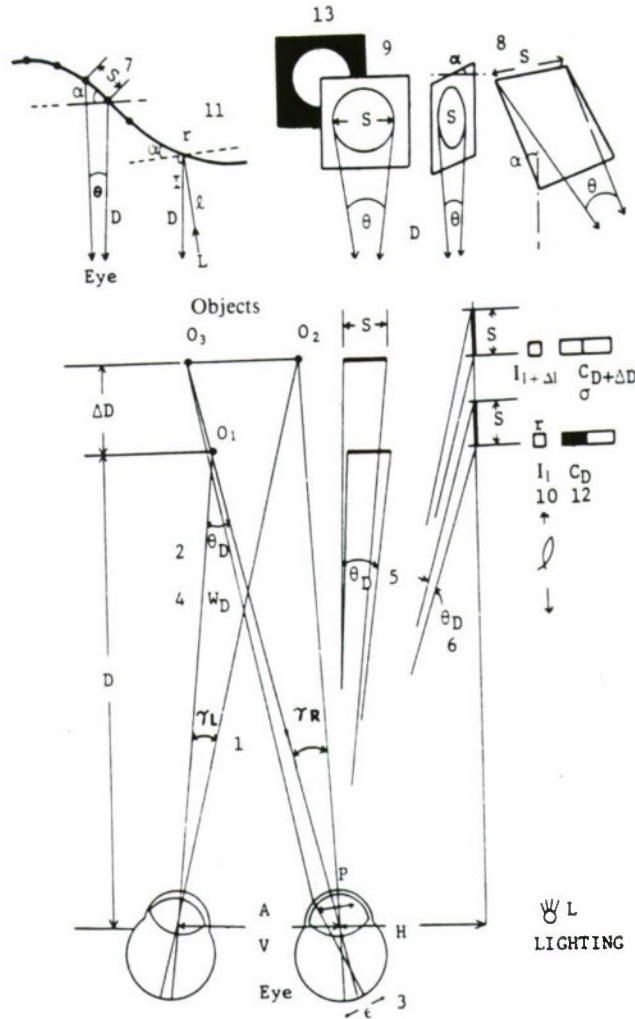
Figure 1.– Illustration of visual cues for depth perception.

1: Binocular parallax $\gamma_L - \gamma_R = \theta_D - \theta_{D+\Delta D}$ at the distance A between pupils.
2: Convergence cue $\theta_D - \theta_{D+\Delta D}$.
3: Blurring cue $\varepsilon$ of accommodation on pupil diameter P.
4: Motion parallax $\gamma_L - \gamma_R$ or $\omega_D - \omega_{D+\Delta D}$ at monocular moving vision of distance A or speed V.
5: Transversal size cue $\theta_D - \theta_{D+\Delta D}$.
6: Longitudinal size cue on depth direction axis at distance H.
7: Density cue $[(S/D)\cos\alpha]^{-1}$ of texture on surface at slant $\alpha$.
8: Shape cue at slant.
9: Intersection cue.
10: Brightness cue $I_1 - I_{1+\Delta 1}$, $I = r \cdot L/I^2$ of the object with refractory factor r at lighting distance l under lighting L.
11: Shade cue $I \cos\alpha$ on slanted surface.
12: Air-perspective contrast cue $C_D - C_{D+\Delta D}$ of air scattering constant $\sigma$.
13: Color effect.

Figure 2.– Apparatus for measuring depth thresholds.



Figure 3.– Depth sensitivities of various cues for visual depth perception as a function of viewing distance. Symbols (o,■,△) indicate the averages of five measurements of subject SN and bars on the symbol indicate standard deviations.

| | | |
|---|---|---|
| Binocular parallax: | A | $= 0.065$ m, $\Delta\theta = 25''$ |
| Motion parallax: | $V_{max}$ | $= 0.8$ m/sec, $\Delta\omega = 4'$/sec, $\omega a = 6°$/sec |
| Accommodation: | P | $= 0.005$ m of the natural pupil, $\Delta\theta_A = [1/1.2]'$ |
| Air-perspective: | $C_0$ | $= 1, \sigma = 1$ km, $\Delta C = 11\%$ of $C_D$ [$\pm 1$ dB], $C_{min} = 0.02$ |
| Transversal size: | $\Delta\theta_s$ | $= 2.5\%$ of retinal size $\theta_s$ |
| Texture/Longitudinal size: | $\Delta\theta_s$ | $= 2.5\%$ of retinal size $\theta_s$ |
| Convergence: | $\Delta\theta_s$ | $= 10$ min |
| Brightness: | $I/\Delta I$ | $= 0.02$ |

Figure 4.– Thresholds of binocular parallax as a function of viewing distance.
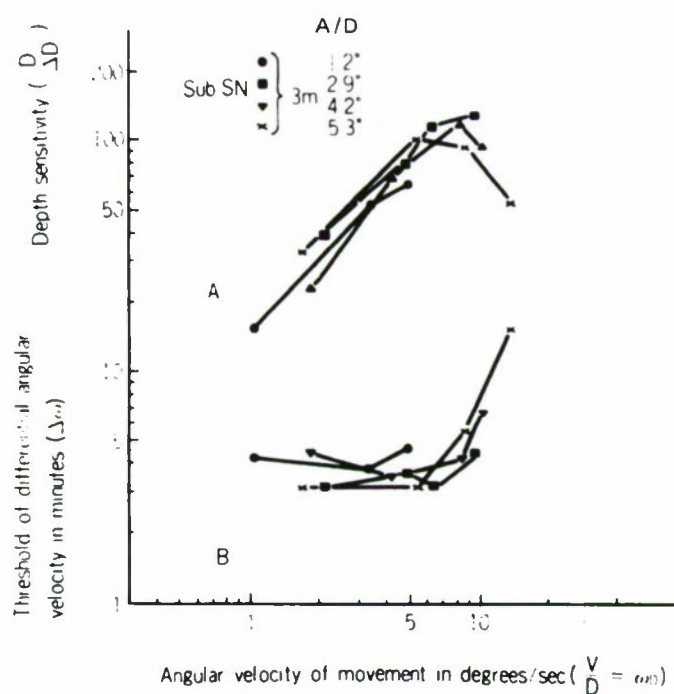


Figure 5.– Depth sensitivities D/ΔD (curves of A) and the threshold of parallax velocity Δω (curves of B) as functions of angular velocity of movement ωD = V/D and movement distance A at a viewing distance of 3 m.
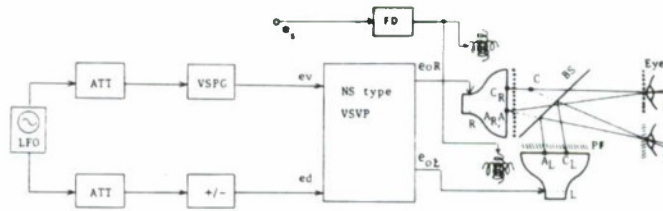
Figure 6.– Diagram of experiments for binocular parallax and size changing cue. LFO:low frequency oscillator. ATT:attenuator. VSPG:variable size square pattern generator. VSVP:versatile stereoscopic video processor. FD:fixed delay. $e_v$:television video signal of original picture. $e_0$:phase-modulated video signal for left or right eye. $e_d$:depth signal for modulation of binocular parallax. $e_s$:synchronous signal. BS:beam splitter. PF:polarizing filter. $A_{R,L}$; $C_{R,L}$:position of pictures. A;C:perceived position.
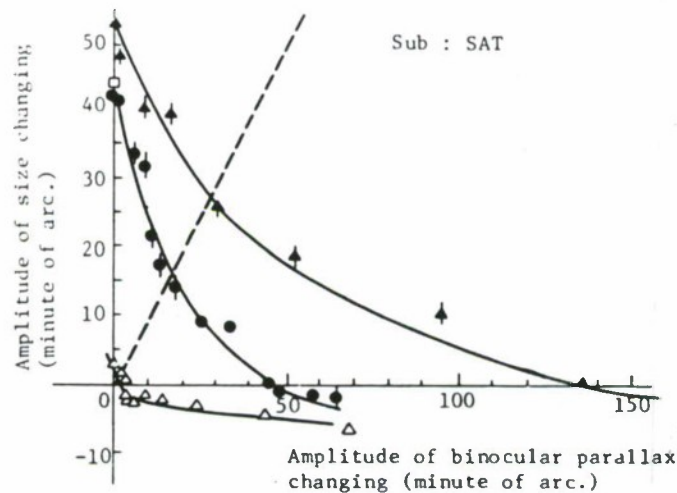


Figure 7.– Interactive effects of depth sensation from two kinds of cue, binocular parallax and changing size cue with oscillating amplitude.

$\Delta$: conditions for the threshold of depth motion perception

$\bullet$,$\blacktriangle$: conditions for equal depth sensation at two levels of suprathreshold

$\square$: condition of only size cue in monocular vision for equal depth sensation with that of $\blacktriangle$

--- : condition in actual moving

Sine-wave oscillation frequency, 1 Hz. Middle size, 6.4 cm (2.71°) x 6.4 cm. Back luminance, 1 cd/m$^2$; Pattern, 30 cd/m$^2$. Viewing distance, 1.35 m.
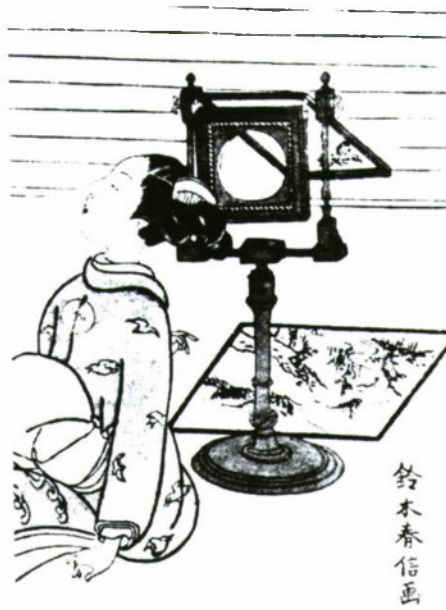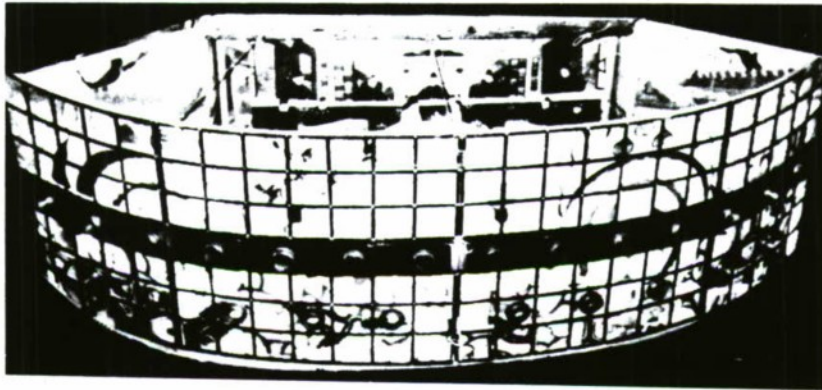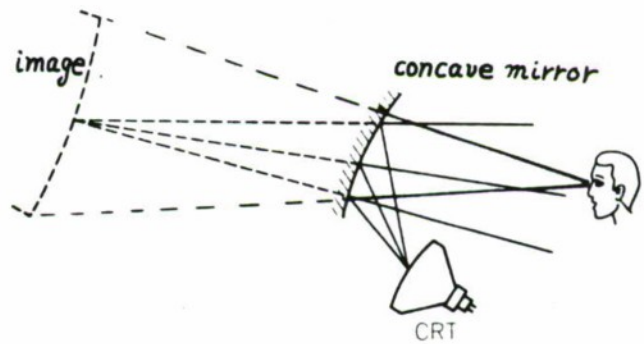
Figure 8.– Old viewing device called vue d optique (nozoki-karakuri) with one lens-mirror (a) and with 24 lenses (b) in Japan and same type (c) in China and one kind (ref. 23) of new wide television system (d).

# PICTURE PERCEPTION:  OTHER CUES

# THE EYES PREFER REAL IMAGES

Stanley N. Roscoe
ILLIANA Aviation Sciences Limited
Las Cruces, New Mexico

For better or worse, virtual imaging displays are with us in the form of narrow-angle combining-glass presentations, head-up displays (HUD), and head-mounted projections of wide-angle sensor-generated or computer-animated imagery (HMD). All of our military and civil aviation services and a large number of aerospace companies are involved in one way or another in a frantic competition to develop the best virtual imaging display system. The success or failure of major weapon systems hangs in the balance, and billions of dollars in potential business are at stake. Because of the degree to which our national defense is committed to the perfection of virtual imaging displays, a brief consideration of their status, an investigation and analysis of their problems, and a search for realistic alternatives are long overdue.

## CURRENT STATUS

All of our currently operational tactical fighter aircraft are equipped with HUDs. Helicopters are navigated and controlled, and their weapons are delivered, with a variety of imaging displays including, in addition to HUDs, both panel-mounted and head-mounted image intensifiers and forward-looking infrared (FLIR) and low-light TV displays. Even some strategic aircraft and a few commercial airliners contain virtual imaging displays. A new generation of remotely piloted vehicles (RPV) are intended to be flown by reference to wide-angle but relatively low-resolution sensor imagery presented stereoscopically by head-mounted binocular displays. And Detroit is about to offer HUDs for cars.

## THE TROUBLE WITH HUDS AND HMDS

As for the operational problems, about 30% of tactical pilots report that using a HUD tends to cause disorientation, especially when flying in and out of clouds (Barnette, 1976; Newman, 1980). Pilots frequently experience confusion in trying to maintain aircraft attitude by reference to the HUD's artificial horizon and "pitch-ladder" symbology, particularly at night and over water, and there are documented cases of airplanes becoming inverted without the pilots' awareness (Kehoe, 1985). Pilots have also reported a tendency to focus on the HUD combining glass instead of the outside real-world scene (Jarvi, 1981; Norton, 1981). The resulting myopia is a special case of the more general anomaly known as "instrument myopia" (Hennessy, 1975).

### Misaccommodation of the Eyes

Whatever the cause, it is a repeatedly observed experimental fact that our eyes do not automatically focus at optical infinity when viewing collimated virtual images, but lapse inward toward their dark focus, or resting accommodation distance, at about arm's length on average (Hull, Gill, and

Roscoe, 1982; Iavecchia, Iavecchia, and Roscoe, 1988; Norman and Ehrlich, 1986; Randle, Roscoe, and Petitt, 1980). The perceptual consequence of positive misaccommodation is that the whole visual scene shrinks in apparent angular size. This shrunken appearance causes distant objects to be judged farther away than they are, and anything below the line of sight, such as the surface of the terrain or an airport runway, appears higher than it really is relative to the horizon (Roscoe, 1984, 1985).

The effect of the HUD optics is illustrated in figure 1. The experiment was conducted by Joyce and Helene Iavecchia at the Naval Air Development Center in Pennyslvania. A HUD was set up on one rooftop and a "scoreboard" assembly with selectively lighted numerals of various sizes was mounted on top of another building 182 m away and of about the same height. Observers were asked to read scoreboard numbers as they appeared and also numbers presented by the HUD on half the trials. Concurrently, the eye accommodation of the observers was measured with a polar-ized vernier optometer.

Figure 1 shows the average focal responses to the scoreboard numerals and the background terrain beyond the scoreboard, with the HUD turned Off and with it turned On. In either case the observers' focal responses were highly dependent on their individual dark focus distances; in fact, knowing each individual's dark focus accounted for 88% of the variance in focal responses under all conditions of the experiment. Excluding Observer 9, whose dark focus was almost three diopters (D) beyond infinity, the average for the remaining nine emmetropes was 1.06 D, or just short of 1 m.

But the striking result shown in figure 1 is the fact that when the HUD was turned On, for all 10 observers, focus shifted inward from an average of 0.02 D, or 50 m, to an average of 0.20 D, or 5 m. Once again excluding Observer 9, the average inward shift was from 0.27 D, about 4 m, to 0.47 D, about 2 m. Although such shifts have little effect on the apparent clarity of the visual scene, they have tremendous effects on the apparent size, distance, and angular direction of terrain features.

## Accommodation and Apparent Size

Despite wide individual differences among observers, the average apparent size of objects is almost perfectly correlated ($r > 0.9$) with the distance at which the eyes are focused (Benel, 1979; Hull, Gill, and Roscoe, 1982; Iavecchia, Iavecchia, and Roscoe, 1983; Roscoe, Olzak, and Randle, 1976; Simonelli, 1979). Thus, the positive misaccommodation induced by collimated HUD symbology can partially account for the fact that pilots flying airplanes or flight simulators by reference to virtual imaging systems make fast approaches, round out high, and land long and hard (Campbell, McEachern, and Marg, 1955; Palmer and Cronn, 1973).

Such biased judgments also partially account for the fact that helicopter pilots flying with imaging displays frequently collide with trees and other surface objects and the fact that the U. S. Air Force between 1980 and 1985 lost 73 airplanes in clear weather because of pilot misorienta-tion, resulting in controlled flight into the terrain (54), or disorientation resulting in loss of control (19) while flying by reference to collimated HUDs (Morphew, 1985). When flying by reference to panel-mounted or head-mounted imaging displays, helicopter pilots approach objects slowly and tentatively, and still they are frequently surprised when an apparently distant tree or rock suddenly fills the wide-angle sensor's entire field of view.

Fixed-wing airplane pilots flying with HUDs also judge a target to be farther away and the dive angle shallower than they are, resulting in almost-always-fatal "controlled-flight-into-the-terrain" accidents. In the U.S. Air Force, such accidents have continued to occur at the rate of about one per month since HUDs came into general use at the beginning of this decade. Two months ago (June 1987) an F-16 left a smoking hole in the ground, and last month it was an F-111. The Navy's experience has been essentially the same.

## Optical Minification

Misorientation and disorientation with panel-mounted and some head-mounted imaging displays are exacerbated by the fact that limited display size and the need to display the widest practical outside visual angles typically result in drastic optical minification, which adds to the perceptual minification caused by the misaccommodation. If the display area were not so limited and could be varied to accommodate the wide individual differences in dark focus distances, images of the outside world could be magnified by appropriate amounts to neutralize each individual's perceptual bias. The average magnification required would be X1.25 (Roscoe, 1984; Roscoe, Hasler, and Dougherty, 1966), but this value would be correct for only a portion of the population, possibly requiring stricter pilot selection.

## Image Quality

Display minification and perceptual biases are two sources of error in human judgments of size, distance, and angular location, but there are other sources of error as well, namely, the variable errors associated with adverse ambient viewing conditions (atmospheric attenuation and reduced illumination), the limited resolution of cameras and display devices, and the further loss of resolution with image intensification. All of these factors serve to reduce contrast and detail, the principal components of image quality, and the accuracy with which people can extract positions, rates, and accelerations relative to outside objects in the visual environment.

## DISPLAY ALTERNATIVES

Because of the adverse effects of virtual images on eye accommodation, as well as the optical minification and poor image quality typically associated with sensor-generated displays, our judgments of spatial relations are simply not good enough to support complex flight missions as safely or effectively as we need. To date the advocates of virtual image displays have adamantly refused to acknowledge the implication of misaccommodation in the misorientation and disorientation of pilots flying with HUDs. Instead they have attributed the problems primarily to the limited fields of view afforded by the combining glasses used with current systems.

To address the limited-field-of-view problem, each of our military services, including the Marines, is spending millions of dollars a year—to say nothing of the IR&D funds invested by private companies—to develop wide-angle, head-mounted imaging displays, in many cases coupling camera line-of-sight to head or eye orientation. Still clinging to the assumption that the eyes will focus collimated images at optical infinity, the advocates of head-mounted displays and

head-coupled sensors now promise that a pilot will be able to maintain geographic orientation and make veridical judgments of distances, rates of closure, and angular directions to visible navigation points and targets.

To dispel any doubt that such promises will come true, designers of some sensor and display systems are delivering imagery from two cameras independently to the two eyes to provide stereoscopic viewing (or even hyperstereo by exaggerating the interocular distance between the cameras). Many are convinced that stereo viewing will create an illusion of "remote presence" and thereby improve judgments of size, distance, and angular location sufficiently to make it unnecessary to provide automatic sensors of vehicle positions and rates for navigation and obstacle avoidance. Experience with head-mounted displays, whether binocular or biocular (both eyes receiving the same images), does not warrant these wishful thoughts.

Evidence from a variety of experimental and operational contexts indicates that binocular judgments of size and distance are not markedly better than monocular judgments, except at very short distances (as in threading a needle). In fact, Holway and Boring (1941) found monocular size judgments to be more nearly veridical than binocular judgments when good distances cues are present. In any case, the large bias errors in size, distance, and angular position judgments caused by misaccommodation to virtual images would more than cancel any minor benefits of disparate images to the two eyes.

In the absence of some striking breakthrough in human genetic engineering, the long-range prognosis for head-mounted displays is not good. Not only do our eyes refuse to behave as display designers would like to believe, but the illusion of vection induced by the "streaming" of objects near the periphery of wide-angle views often leads to motion sickness, particularly with head-coupled sensors and the consequent smearing of the images with head movements. Unfortunately our sole dependence on virtual imaging displays for tactical missions (HUDs now and HMDs in the future) has resulted in almost total suppression of research and development of more easily optimized direct-view displays of sufficient angular size to provide the needed fields of view with appropriate magnification.

## WHAT CAN BE DONE

If we dismiss the genetic engineering approach, there are still several reasonable courses of action. In the short run, these include (1) trying to "fix" the HUD optics to compensate for the misaccommodation that leads to misorientation, and (2) modifying the ambiguous HUD symbology that leads to attitude reversals and subsequent disorientation. In the longer run, abandon the virtual image approach and concentrate on large, integrated forward-looking and downward-looking direct-view displays in which computer-animated flight attitude, guidance, and prediction symbology is superposed on sensor-generated real-world imagery.

### Fixing the HUD

To induce pilots to focus at optical infinity when viewing virtual images, Norman and Ehrlich (1986) in Israel introduced a negative focal demand of -0.5 D with the desired result, although there were wide individual differences in responses as a function of individual dark-focus

distances. Thus, the first experimental fix should be the addition of variable optical refraction to offset each individual pilot's inward focal lapse induced by the HUD's virtual images. Turning the HUD On would require a key coded to select the pilot's specific correction based on the dark focus. At this time, no one can be sure how successful this fix will be, but it must be tried.

Almost as important is the complete redesign of HUD symbology. Just how complicated and confusing it is can be appreciated from the estimate of an Army Instructor Pilot that an average student helicopter pilot requires 200 hr of simulator and flight training to master the gaggle of symbols (personal communication). Furthermore, the attitude presentation in fixed-wing airplanes is conducive to horizon and pitch-ladder control reversals that result in disorientation and "graveyard spirals" at night and in marginal weather. At the very least, a frequency-separated predicted flightpath "airplane" symbol that banks and translates in immediate response and in the same direction as control inputs should replace the present velocity vector and acceleration symbology (Roscoe, 1980, Ch. 7; Roscoe and Jensen, 1981).

## Presenting the Big Picture

If head-mounted, wide-angle imaging displays are ever to be safe and successful, the apparent minification of the outside world will have to be compensated for by individually selectable optical magnification, or the eyes will have to be induced to focus at or near optical infinity, as in the case of HUDs. Neither approach will be simple. Furthermore, the whole virtual image display concept depends on a gross reduction, rather than any increase, in the weight of any head-mounted device to be used in a high-G environment. All things considered, it is surely premature to give up on direct-view, panel-mounted displays.

Large, integrated, direct-view displays offer many advantages in terms of visual performance as well as ease of achievement and lower cost. Eyes focus real images more accurately than virtual images (Hull, Gill, and Roscoe, 1982; Iavecchia, Iavecchia, and Roscoe, 1988; Randle, Roscoe, and Petitt, 1980). Although many with 20/20 vision cannot focus out to optical infinity, all emmetropes can focus at the distance of cockpit instrument panels. Thus, although magnification of sensor-generated or computer-animated images of the outside world will be required, as it is with direct-view projection periscopes (Roscoe, 1984; Roscoe, Hasler, and Dougherty, 1966), a single, fixed-magnification factor of about X1.25 will suffice for most emmetropes.

To make room for large forward-looking and downward-looking (and possibly sideways-looking) displays, a lot of single-variable dedicated instruments and controls will have to be replaced by insets that appear selectively on the large displays as a function of the mission phase, aircraft configuration, mode of operation, weather and traffic, system malfunctions, and in the case of military aircraft, weapon selection. Furthermore, with the ever-increasing complexity of aircraft systems and military missions, many future airplanes—despite their high degrees of automation—will require at least two pilots with a redistribution of functions and available information.

In the military there will always be a heavy premium on being able to take advantage of whatever is visible to the naked eye. However, trying to combine synthetic imagery with contact visibility compromises both, and a strong case can be made for distributing operational functions and information sources between an "inside" pilot and an "outside" pilot. The inside pilot would normally do all the flying in instrument meteorological conditions (IMC) and most of the flying under visual meteorological conditions (VMC), using a direct-view, wide-angle projection periscope and

the large, panel-mounted pictorial displays surrounding the pilot deep inside the airplane. The outside pilot would use his or her eyes to supplement the imaging sensors, do most of the communicating and procedural housekeeping, and fly any maneuver that requires direct contact visibility.

# REFERENCES

Barnett, J. F.: Role of Head-up Display in Instrument Flight. IFC-LR-76-2, Randolph Air Force Base, TX: Instrument Flight Center, 1976.

Benel, R. A.: Visual Accommodation, the Mandelbaum Effect, and Apparent Size. BEL-79-1/ AFOSR-79-5, New Mexico State Univ., Behavioral Engineering Laboratory, Las Cruces, 1979. (Dissertation Abst. Intern., vol. 40., no. 10B, 1980, pp. 5044; Univ. Microfilm No. 80-08974.)

Campbell, C. J.; McEachern, L. J.; and Marg, E.: Flight by Periscope. WADC-TR-55-142, Wright-Patterson Air Force Base, OH: Wright Air Development Center, Aero Medical Laboratory, 1955.

Hennessy, R. T.: Instrument Myopia. J. Opt. Soc. Amer., vol. 65, no. 10, 1975, pp. 1114-1120.

Holway, A. H.; and Boring, E. G.: Determinants of Apparent Size with Distance Variant. Amer. J. Psychol., vol. 54, no. 1, 1941, pp. 21-37.

Hull, J. C.; Gill, R. T.; and Roscoe, S. N.: Locus of the Stimulus to Visual Accommodation: Where in the World, or Where in the Eye? Human Factors, vol. 24, no. 3, 1982, pp. 311-319.

Iavecchia, J. H.; Iavecchia, H. P.; and Roscoe, S. N.: The Moon Illusion Revisited. Aviation, Space, Environ. Med., vol. 54, no. 1, 1983, pp. 39-46.

Iavecchia, J. H.; Iavecchia, H. P.; and Roscoe, S. N.: Eye Accommodation to Head-up Virtual Images. Human Factors, vol. 30, no. 6, 1988, pp. 689-702.

Jarvi, D. W.: Investigation of Spatial Disorientation of F-15 Eagle Pilots. ASD-TR-81-5016, Wright-Patterson Air Force Base, OH: Aeronautical Systems Division, 1981.

Kehoe, N. B.: Colonel Kehoe's Spatial Disorientation (SDO) Incident in a F-15 during VMC. In G. B. McNaughton (Ed.), Aircraft Attitude Awareness Workshop Proceedings, Wright-Patterson Air Force Base, OH: Flight Dynamics Laboratory, 1985, pp. 1-5-1 to 1-5-4.

Morphew, G. R.: Transcript of Open Forum Session. In G. B. McNaughton (Ed.), Aircraft Attitude Awareness Workshop Proceedings, Wright-Patterson Air Force Base, OH: Flight Dynamics Laboratory, 1985, p. 3-8-1.

Newman, R. L.: Operational Problems with Head-up Displays During Instrument Flight. AFAMRL-TR-80-116, Wright-Patterson Air Force Base, OH: USAF Aerospace Medical Research Laboratory, 1980.

Norman, J.; and Ehrlich, S.: Visual Accommodation and Virtual Image Displays; Target Detection and Recognition. Human Factors, vol. 28, no. 1, 1986, pp. 135-151.

Norton, P. S. (Moderator) and members of the SETP Cockpit Design Subcommittee: Proc. Soc. Exp. Test Pilots Aviation Safety Workshop, New York: AIAA, 1981, pp. 19-47.

Palmer, E.; and Cronn, F. W.: Touchdown Performance with a Computer Graphics Night Visual Attachment. Proc. AIAA Visual and Motion Simulation Conference (AIAA paper 73-927), New York: AIAA, 1973.

Randle, R. J.; Roscoe, S. N.; and Petitt, J.: Effects of Magnification and Visual Accommodation on Aimpoint Estimation in Simulated Landings with Real and Virtual Image Displays. NASA TP-1635, 1980.

Roscoe, S. N.: Aviation Psychology. Iowa State University Press, Ames, 1980.

Roscoe, S. N.: Judgments of Size and Distance with Imaging Displays. Human Factors, vol. 26, no. 6, 1984, pp. 617-629.

Roscoe, S. N.: Bigness is in the Eye of the Beholder. Human Factors, vol. 27, no. 6, 1985, pp. 615-636.

Roscoe, S. N.; Hasler, S. G.; and Dougherty, D. J.: Flight by Periscope: Making Takeoffs and Landings; the Influence of Image Magnification, Practice, and Various Conditions of Flight. Human Factors, vol. 8, no. 1, 1966, pp. 13-40.

Roscoe, S. N.; Olzak, L. A.; and Randle, R. J.: Ground-referenced Visual Orientation with Imaging Displays: Monocular versus Binocular Accommodation and Judgments of Relative Size. Proc. AGARD Conf. Visual Presentation of Cockpit Information Including Special Devices for Particular Conditions of Flying, Neuilly-sur-Seine, France: NATO, 1976, pp. A5.1-A5.9.

Roscoe, S. N.; and Jensen, R. S.: Computer-Animated Predictive Displays for Microwave Landing Approaches. IEEE Trans. Syst. Man, and Cybernet., vol. SMC-11, no. 11, 1981, pp. 760-765.

Simonelli, N. M.: The Dark Focus of Accommodation: Its Existence, its Measurement, its Effects, BEL-79-3/AFOSR-79-7, New Mexico State Univ., Behavioral Engineering Laboratory, Las Cruces, 1979. (Dissertation Abstr. Intern., vol. 41, no. 02B, 1980, p. 722; Univer. Microfilms No. 80-17984).
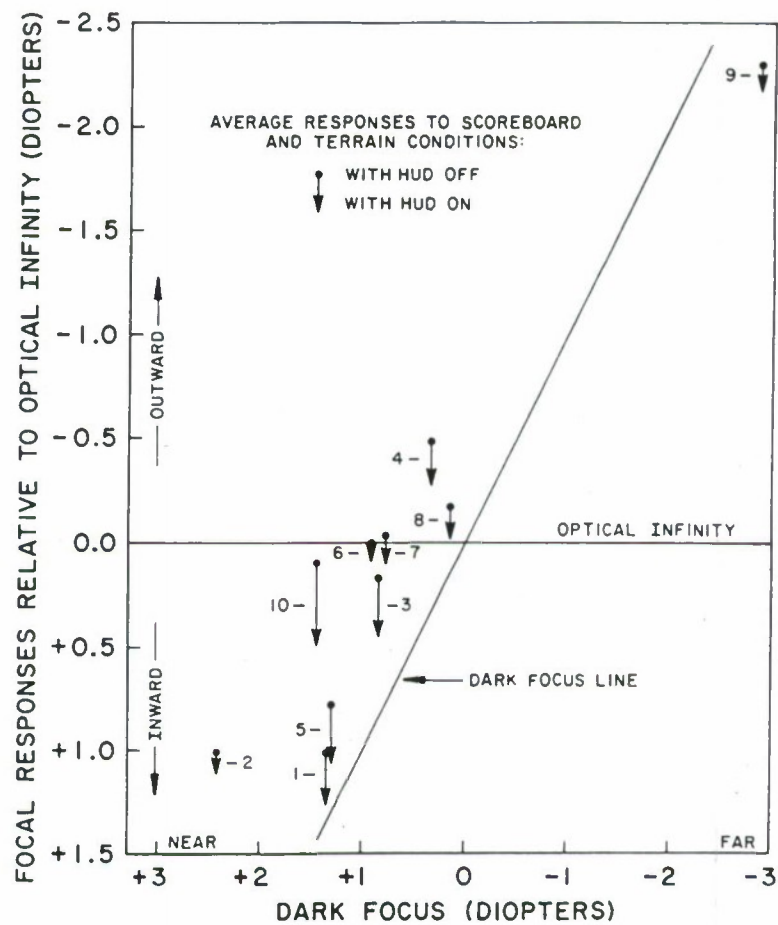
Figure 1.– Average focal responses to the scoreboard and the terrain conditions with HUD On and Off, plotted against each individual's dark focus.

# SPATIAL ISSUES IN USER INTERFACE DESIGN FROM A GRAPHIC DESIGN PERSPECTIVE

Aaron Marcus
Aaron Marcus and Associates
1196 Euclid Avenue
Berkeley, California 94708-1640

## INTRODUCTION

The user interface of a computer system is a visual display that provides information about the status of operations on data within the computer and control options available to the user that enable adjustments to these operations. From the very beginning of computer technology the user interface was a spatial display, although its spatial features were not necessarily complex or explicitly recognized by the users. All text and nonverbal signs appeared in a virtual space generally thought of as a single flat plane of symbols.

Current technology of high-performance workstations permits any element of the display to appear as dynamic, multicolor, three-dimensional signs in a virtual three-dimensional space. The complexity of appearance and the user's interaction with the display provide significant challenges to the graphic designer of current and future user interfaces. In particular, spatial depiction provides many opportunities for effective communication of objects, structures, processes, navigation, selection, and manipulation. The following discussion presents issues that are relevant to the graphic designer seeking to optimize the user interface's spatial attributes for effective visual communication.

## CURRENT SPATIAL APPROACHES TO USER INTERFACE DESIGN

In all user interfaces, there is a need to present data objects, processes, their status, and structures of various kinds. In addition, the designer of a user interface must determine means for enabling the user to navigate among these objects, to select them, and to manipulate them in various ways. Influenced by the introduction of the Xerox Star and Apple Macintosh computers in the early 1980s, computer graphics programmers have emphasized recently the multiwindowed desktop metaphor as a basis for appearance and interaction.

The desktop spatial metaphor assumes that the viewer is looking at a flat background, with one or more rectangular windows in front of (or on top of, according to the implied orientation of the conventional horizontal desktop) the background plane. The windows may tile the foreground or may overlap in various ways. Icons, or other small signs, standing for objects, processes, structures, or data, can appear in the background plane or in the window planes. In addition to windows, various menus and dialogue boxes can appear within windows or in front of any or all the windows. In front of all of these elements, cursors may float across the visual field. Any of the windows or the background may contain graphics images that depict a deep three-dimensional space.

The space is designed as a shallow layering of foreground, middle ground, and background, reminiscent of traditional shallow spatial compositions in modern painting (Loran, 1963; Berkman, 1949). This multiple-layered composition is also reminiscent of layered cartoon animation cells, a kind of two-and-one-half-dimensional space, as it is sometimes called.

Certain visual enhancements to the depiction of objects in the space are typically used to help the viewer understand the spatial composition. These include the following techniques: (1) drop shadows, (2) beveled edges, (3) highlighting and lowlighting, and (4) shrinking and growing.

For example, drop shadows, typically directed to the lower right, help to convey the layering of windows, pull-down or pop-up (more explicitly, pop-in-front-of) menus, or dialogue boxes. In some user interfaces, icons, buttons, switches, menu elements, or entire rectangles of menus, dialogue boxes, or windows, may be given beveled sides so that they appear to protrude toward the viewer. Sometimes their sides are colored with varying levels of gray-value to strengthen the illusion of three-dimensional form and a light source, often implied to be located at the upper left. In addition, entire windows or other areas of the screen may be highlighted to come forward to the viewer, while other windows may be lowlighted to suggest that they are farther back in space. Elements sometimes change their size and appearance; for example, an icon may enlarge to become a window. This is often shown as a spatial growth in two dimensions, which contributes to the illusion of overlapping elements.

These techniques are similar to those employed by designers to enhance information-oriented graphics, such as the design of charts, maps, and diagrams (Herdeg, 1981). They have distinct communication value from a graphic design point of view. These spatial qualities accomplish the following:

1. Distinguish various elements on the screen

2. Help the viewer to recognize particular classes of objects

3. Add charm or appeal to the design style of the user interface

4. Convey corporate or product design conventions

Besides the traditional desktop, the image of the control panel is also used in some user interfaces, in which part or all of the screen may convey one or more flat panels with switches, knobs, and other control devices. A variant on the desktop is the giant desktop in which the viewer sees one part of the background through a viewport and must use scrolling devices to examine other areas. Another variant of the desktop might be called the multiple desktop in which the viewer may move from desktop to desktop by zooming, sudden cuts or pops, or other visual techniques. A memorable approach using sound cues to aid spatial cues was presented by the MIT Architecture Machine Group's spatial data management system (Bolt) in the 1970s in which the background plane zoomed toward the viewer with an audible whoosh as the viewer suddenly dropped onto a layer below with an audible popping sound. Apple's Hypercard and similar hypertext products generally extend the notion of the screen as a set of planes.

## OTHER SPATIAL METAPHORS

Programmers have experimented with other spatial metaphors to facilitate human-computer communication. One alternative is the metaphor of architecture. The Learning Company, for example, has offered since the early 1980's an award-winning children's game called Rocky's Boots, programmed by Warren Robinet, that provides the viewer with the cognitive model of a set of rooms, each with entrances and exits. The screen display communicates a set of spaces linked by the topology of familiar architectural experiences. Another approach was taken in the work of Gould and Finzer (1984). They proposed a cognitive model of theater, in which the entire display was depicted as a stage set. This approach implies a deeper spatial metaphor than the traditional desktop.

Other approaches are possible as workstations provide ever greater capabilities to manipulate three-dimensional reality. For example, at the Microcomputer Technology Consortium, Austin, TX, the Semnet project proposed a deep space for viewing and manipulating a semantic network. Another example is the head-mounted display project at NASA Ames Research Center, Moffett Field, CA, begun by Michael McGreevy in which the viewer sees a full three-dimensional environment for all appearance and interaction imagery. With the advent of screens using Adobe's PostScript picture definition language, as in Sun and Next's products, it is possible to display screen metaphors using the building or even the urban environment as a basis for spatial communication of the user interface. All that is required is a set of familiar symbols, a familiar spatial arrangement, and a familiar ritual for interacting with them. Videogames in the entertainment industry have employed routinely a variety of spatial idioms, including rooms, buildings, and landscapes to convey the field of action.

## FUTURE DIRECTIONS

Within the entertainment field and within current user interface design, future directions of spatial representation are already emerging. Two areas of emphasis are depictions of deep space and depictions of three-dimensional objects.

In commercial cable and broadcast television and in the film industry (Morgan and Symmes, 1983), there has been a continuous fascination with depictions of deep space. The title sequence of the *Star Wars* movie, in which text moves backwards at a steep angle from the viewer, inherits a tradition from older films. Today, it is routine for evening news programs, weather reports, movie introductions, and station breaks to feature photographic images, typography, and other elements of flying logos swirling about within deep spatial representations.

All depictions of surfaces, projected light and cast shadows, and dynamic objects in computer graphics are currently very expensive to produce, requiring significant budgets, time, personnel, and equipment. However, the creators of sophisticated animation software, like Wavefront, are broadening the base of hardware and user groups, so that the industry in general will be nurtured with more powerful spatial display and image rendering capabilities. Eventually these capabilities will be routinely available for widespread use in the depiction of user interface components.

Even without expensive workstations, it is possible to display three-dimensional objects as components of the user interface. A current music editing software package on the Commodore Amiga, for example, shows solid pillars and an arch framing the sides and top of the controls for musical composition.

## SPATIAL DEPTH CUES

The use of spatial relations to depict the elements of the user interface suggests that designers may find it useful to review Gibson's list of visual cues that establish the perception of space. These perspective experiences are summarized in Hall's book, *The Hidden Dimension* (1982). Briefly, the taxonomy of spatial depth attributes is the following:

Position
    Texture: gradual increase in density of texture of a receding surface
    Size: gradual decrease in size of distant objects
    Linear perspective: parallel lines receding to vanishing points

Parallax
    Binocular: an image with shifted object locations for each eye
    Motion: objects moving at uniform speeds appear slower if distant

Other Cues
    Aerial perspective: increased haziness and change in color and contrast with distance
    Blur: objects nearer or more distant than the focal plane appear fuzzy
    Vertical location in the visual field: lower part appears nearer, the upper farther
    Shift in double imagery: in distant views, nearer objects have doubling gradient
    Completeness or continuity of outline: nearer objects overlap farther objects
    Shift of light and dark: abrupt changes appear as edges, gradual as roundness

Some, but not all, of these cues are currently employed within user interfaces in order to create convincing spatial scenes. As user interfaces become more visually complex, designers will utilize more of these depth cues and will consequently need to determine user interface spatial-depiction attributes in a systematic manner.

## RELATION TO INDUSTRIAL OR PRODUCT DESIGN

In addition to more complex spatial metrics and spatial metaphors that unite objects in a continuous space (either the familiar Euclidian, the less familiar non-Euclidian, or even strangely warped topologies), increased sophistication of spatial display also means that the individual components of the user interface can take on elaborate internal spatial structures. All of these typical user interface components, such as windows, menus, dialogue boxes, control panels, icons, and cursors, can acquire significant plastic form attributes.

Consider the following examples of possible attribute sets:

Windows with solid extruded shapes for title area and scroll bars

Scroll bars appearing as translucent round columns with the symbol for the visible portion of the screen represented as a solid tube sliding within them

Windows as the front surface of rectangular parallelopipeds, with regular conventions of semantics assigned to the other faces of the solid

Icons as three-dimensional blocks with internal moving parts, whose surface characteristics (metallic, rough, warm, etc.) or interlocking features might contribute to their denotation

Cursors as large, three-dimensional portraits whose pointing fingertips focus the user's attention on a particular screen component while their facial expression conveys important connotative content

At this point, user interface designers would benefit by examining the history and current practice of professionals in graphic design, architecture, industrial design, and product design (Herdeg, 1981; Jencks, 1982; Pevsner, 1963; *Industrial Design Magazine*). In contemporary industrial design, for example, one finds a dialectic taking place between minimalist, Apollonian approaches (International style, Bauhaus style, etc.) in which all objects have a highly consistent, limited selection within attribute space, and the more exuberant, Dionysian approaches (Memphis style, product semantics style, post-modem style) in which eclectic, exotic, wildly different attribute selection reigns. User interface design at this point leaves the engineering domain and enters the world of aesthetic styling, which contributes significantly to the marketing of products worldwide. It is also in this realm of the user interface as plastic, shaped artifact, that corporate design or product design standards influence the three-dimensional attribute selections (Marcus, 1984, 1985).

As user interface design takes on more spatial attributes, the collection of symbols in space take on cultural characteristics far more complicated than the basic issues of ergonomic design. It would seem reasonable for user interface designers to consider the discipline of proxemics (Hall, 1963), the science of interpersonal space, for guidance in user-computer spaces.

## SUMMARY

Aided by advancing technology and spurred both by the need for depicting increasing amounts of data and functions and by market interest, user interface design is taking on more spatial characteristics. User interface graphic designers will need to coordinate, unify, and optimize for communication effectiveness a very broad, deep hierarchy of spatial attributes for every component of the interface. Lessons can be learned by examining the theory and practice of professionals in other disciplines who have also worked with complex spatial structures, both as matters of geometry and as cultural artifacts. The reading list is intended as an initial guide to the literature of these allied disciplines. The scope and rate of change within user interface design promises to offer an exciting opportunity and test of skill for the human mind in shaping three-dimensional forms for pictorial communication.

# READING LIST

Berkman, Aaron, Art and Space, Social Sciences, New York, 1949, 175 pp.

Bolt, Richard A., The Human Interface: Where People and Computers Meet, Lifetime Learning Publications, Wadsworth, Belmont, CA, 113 pp.

Critchlow, Keith, Order in Space: A Design Source Book, Studio Book, Viking Press, New York, 1970, 120 pp.

Ghyka, Matila, The Geometry of Art and Life, Sheed and Ward, New York, 1966, 174 pp.

Gould, Laura, and William Finzer, Programming by Rehearsal, Xerox Palo Alto Research Center Technical Report SCL-84-1, May, 1984, 133 pp (also a shorter version in Byte, Vol. 9, No. 6 June 1984).

Gregory, R. L., The Intelligent Eye, McGraw-Hill, New York, 1970, 191 pp.

Gregory, R. L., Eye and Brain: The Psychology of Seeing, World University Library, McGraw-Hill, New York, 1968, 251 pp.

Hall, Edward T., The Hidden Dimension, Anchor Books, New York, 1982, 217 pp.

Hall, Edward T., The Silent Language, Fawcett, New York, 1963, 192 pp.

Hambidge, Jay, The Elements of Dynamic Symmetry, Dover Publications, New York, 1967, 133 pp.

Herdeg, Walter, Diagrams, Graphis Press, Zurich, 1981, 205 pp.

Industrial Design Magazine, published by Design Publications, Inc., 330 West 42nd Street, New York, NY 10036, monthly.

Ivins, William M., Jr., Art and Geometry: A Study in Space Intuitions, Dover Publications, New York, 1946, 113 pp.

Jencks, Charles and William Chaitkin, Architecture Today, Harry N. Abrams, New York, 1982.

Lakoff, George and Mark Johnson, Metaphors We Live By, The University of Chicago Press, Chicago, 1984, 242 pp.

Loran, Erle, Cézanne's Composition: Analysis of His Form with Diagrams and Photographs of His Motifs, University of California Press, Berkeley, 1963, 143 pp.

March, Lionel and Philip Steadman, The Geometry of Environment: An Introduction to Spatial Organization in Design, RIBA Publications, London, 1971, 360 pp.

Marcus, Aaron, et al., "Screen Design Guidelines," <u>Proceedings</u>, National Computer Graphics Association Annual Conference and Exposition, 14-18 April 1985, pp. 105-137.

Marcus, Aaron "Corporate Identity for Iconic Interface Design: The Graphic Design Perspective," <u>IEEE Computer Graphics and Applications</u>, IEEE Computer Society, New York, December 1984, Vol. 4, No. 7, pp. 24ff.

Morgan, Hal and Dan Symmes, <u>Amazing 3-D</u>, Little, Brown, and Co., Boston, 1983.

Piaget, Jean and Bärbel Inhelder, <u>The Child's Conception of Space</u>, W. W. Norton and Company, New York, 1967, 490 pp.

Pevsner, Nikolaus, <u>An Outline of European Architecture</u>, Penguin Books, Baltimore, 1963, 496 pp.

Stevens, Peter S., <u>Handbook of Regular Patters: An Introduction to Symmetry in Two Dimensions</u>, MIT Press, Cambridge, 1981, 400 pp.

Thomas, Richard K., <u>Three-Dimensional Design: A Cellular Approach</u>, Van Nostrand Reinhold, New York, 1969, 96 pp.

Thompson, D'Arcy, ed. J. T. Bonner, <u>On Growth and Form</u>, Cambridge University Press, Cambridge, 1966, 346 pp.

Weyle, Hermann, <u>Symmetry</u>, Princeton University Press, Princeton, 1966, 168 pp.

# INTERACTIVE DISPLAYS IN MEDICAL ART

Deirdre Alla McConathy and Michael Doyle
University of Illinois, Department of Biomedical Visualization and College of Medicine
Chicago and Urbana, Illinois

Medical illustration is a field of visual communication with a long history. Leonardo DaVinci, inventor, scientist, and illustrator, is perhaps the best known pioneer of medical art, but many other individuals, such as the famous anatomist Vesalius, also contributed to the development of the profession. Understandably, many factors have impacted the field throughout its growth, but the primary goal of a medical artist–to visually explain information about the health sciences–has remained unchanged. Other goals such as marketing and advertising of products are subsidiary to this central objective of presenting educational imagery to health science professionals and patients alike.

Traditional medical illustrations such as the one shown in figure 1 are static, two-dimensional, printed images—highly realistic depictions of the gross morphology of anatomical structures (Netter, 1948; Pernkopf, 1963; The Urban and Schwarzenberg Collection of Medical Illustrations Since 1896, 1977). Coincidental with technological advances in both medicine and image production, however, is the expansion of the role of medical art. Today medicine requires the visualization of structures and processes that have never before been seen. Complex three-dimensional spatial relationships require interpretation from two-dimensional diagnostic imagery. Pictures that move in real time have become clinical and research tools for physicians.

Medical artists are uniquely qualified to plan and produce visual displays for use in health communication. Basic science courses taken within a medical school curriculum prepare them to be content experts. Prerequisite life drawing, painting, color theory, graphic design and other fine art courses, and subsequent graduate coursework including anatomical drawing and surgical illustration imbue artistic skills. Using instructional design theory, artists plan goals and objectives, perform critical analyses of task and learning performance, and evaluate products and procedures. Medical artists are media technologists as well. They must choose from a plethora of media the appropriate mode of presentation for the specific content being represented. The objective in medical art is to incorporate new technologies as both production tools and modes of final presentation. The artists are therefore knowledgeable of a wide variety of media, including printed images in line, continuous tone or color; projection media such as slides, video, film, and animation; computer graphics; and three-dimensional models and simulators.

In addition to formal instruction, medical artists possess those abilities often attributed to the mystical realm of art. Perhaps because of their comprehensive knowledge base relevant to problems of visual representation, for artists an iterative problem-solving process often becomes automatic to the point of appearing to be intuitive. Previsualization of visual solutions by the artist allows exploration to occur in an effective, if not well-understood, manner. For example, Ansel Adams, renowned for his development of the zone system in black and white photography, was consciously aware of the limitations of film for representing the range of values we are able to see with the human eye. He could, however, mentally image how a landscape would be recorded by film, and thereby "see" a predictable translation to guide him. In a similar manner, medical illustrators use a combination of factual, theoretical and artistic knowledge to previsualize.

Clients and content experts need to be involved in the process of preparing visuals, but many important production decisions pertaining to the final appearance of the image are solely the domain of the artist. Artists are able to identify and manipulate many variables with predictable results and recognize the contributions of unpredictable "happy accidents."

The most fundamental decisions upon initiating a drawing involve characteristics of the light source portrayed. The importance of direction of a light source is well documented. Perceptual psychologists have demonstrated that an upper-left light source is generally the default assumption for a viewer, but direction is only one variable to be considered. Two other important considerations are color temperature and intensity, as each of these convey information about spatial relationships and can be used to invoke affective reactions. The artist sometimes needs to invent the light source, creating an unreality that is more effective than reality. For example, operating room lights provide very diffuse, even lighting of the surgical field to avoid fatigue to the surgeon's eyes; therefore photographs appear to be flat spatially. Surgical illustrators enhance the impression of space by creating an imaginary, directional light source, with strong highlights and cast shadows. Many other artistic decisions, such as viewer station point, composition, and color harmony, all impact the final results, and should be entrusted to professional communicators and qualified artists.

The medical artist embodies a link between the technical and aesthetic realms of visual communication. The skills exemplified by medical artists for the health sciences community can demonstrate an appropriate model for other fields that need to make judgments about visuals from a holistic viewpoint.

The importance of a qualified consultant and producer of visuals cannot be overemphasized. In their report to the National Science Foundation ("Visualization in Scientific Computing," McCormick, DeFanti and Brown (1987)) comment that "Because of inadequate visualization tools, users from industry, universities, medicine, and government are largely unable to comprehend the flood of data produced by contemporary sources such as supercomputers, satellites, spacecraft, and medical scanners. Today's data sources are such fire hoses of information that all we can do is warehouse the numbers they generate, and there is every indication that the number of sources will multiply." The authors suggest that interactive graphics are the best available solution to managing this information deluge. They go on to recommend that interdisciplinary teams of computer scientists, engineers, cognitive scientists, systems support personnel, and artists be enlisted to attack the visualization challenge.

One inevitable question for all types of pictorial displays is how realistic should the image be? Much debate exists as to the appropriate amount of realism necessary to include in different types of visuals. Research of the realism continuum and its effect on learning has not, however, established usable guidelines to be implemented. The current trend in educational resources is toward editing of information within pictures to a more diagrammatic style, whereas efforts to improve simulators are toward maximizing realism. Interactive displays may prove to be a reasonable solution to the editing question by providing users the flexibility of controlling the variable of realism and detail themselves. In reality, however, the issue of optimal levels of detail to include in a particular illustration is most often settled by budgetary constraints or subjective client preferences.

Medical illustrators are involved with the development of interactive visual displays for three different, but not discrete, functions: as educational materials, as clinical and research tools, and as databases of standard imagery used to produce visuals.

Health education visuals are required for a diverse audience including patients, medical students in training, and experienced surgeons. The information depicted may be factual, theoretical, abstract, or motor-skill training. Patient simulators are, for example, important methods for training manual skills because they offer the greatest breadth of learning experience with no risk of damage or discomfort to the patient. A successful simulator should provide a high degree of procedural realism. A three-dimensional model (fig. 2) used to train personnel in the procedure for fetal monitoring exemplifies a traditional type of interactive teaching display.

Monitoring a fetal heart rate during labor requires the insertion of an intra-uterine pressure catheter and the attachment of a scalp electrode to the baby. Placement of the instruments is critical since misapplication can result in devastating damage to the newborn. Correct positioning of the instruments requires the technician to palpate anatomical landmarks and visualize spatial relationships.

To satisfy these requirements in the simulator, medical sculptor Ray Evenhouse mimics soft and bony tissues with layers of synthetic materials. The structures are made from casts of bones and sculptures of soft tissues based on morphometric data. The completed simulator consists of a fetal head that is positioned within the maternal torso by an instructor in a variety of presentations. Visual and tactile realism is essential so that underlying structures such as the anterior and posterior fontanelles and facial features can be palpated to orient the trainee. In addition, the motivational factor induced by a highly aesthetic simulator contributes to the overall success of the model (Evenhouse and McConathy, 1989).

A quite different simulation is represented by an electronic textbook recently developed by Doyle (O'Morchoe and O'Morchoe, 1987) as a tool for teaching histology, the study of cell and tissue biology, to medical students (fig. 3). This prototype system operates on an IBM PC microcomputer fitted with both a high-resolution graphics display, capable of 256 on-screen colors, and a separate monochrome text display. The textbook uses the interactive digital video (IDV) interface, a device-independent process for user interaction with digital video images.

A student operating the system is presented with a menu from which a particular histological section is chosen for viewing. A realistic video image of that section is then called up from the disk and displayed on the color monitor. The student is then able to interact directly with the video image by pointing to an area of interest with a mouse and pressing a button. The system responds by displaying descriptive test on the monochrome monitor, which explains the pertinent facts about that particular image feature. For example, if the student points to a muscle cell within an image of heart tissue, the text which is displayed on the monochrome monitor explains in detail the salient morphological characteristics of cardiac muscle, how this type of muscle tissue compares to skeletal and smooth muscle, and so on. This atlas is an attempt to create an entirely intuitive user interface for the student. A specific goal was to eliminate the distraction of labeling every important histological structure on the screen simultaneously while still allowing the user instant access to the exact conceptual elaboration which he or she desires.

It is possible to reverse the above-mentioned situation so that the student can type in a structure name on the keyboard and the system then displays an image with the pertinent structure

highlighted. The IDV interface can also be used to correlate one image to another so that the selection of a histological structure results in the display of either higher or lower magnification views of that particular image. This allows the student to zoom in from an orienting, low-magnification, light microscopic view, through consecutive higher-magnification images, all the way to the electron microscopic level and back again. The possibility is also being investigated of using a speech synthesizer for text output as well as a voice-recognition system for user text queries.

These examples highlight the range of possibilities for teaching with interactive visuals. The opportunities for students to learn in real time, encounter variations, self-edit information, and adopt learning strategies best suited to their own needs represent a major advancement in education.

Another burgeoning area of interactive displays involves visuals as clinical and research tools. The advent of computer technology in combination with new technologies of diagnostic imaging has provided physicians and researchers new methods of visualization.

The imaging modalities of computed transmission and emission tomography, magnetic resonance imaging and ultrasound are revolutionizing medicine. "Improved 3D visualization techniques are essential for the comprehension of complex spatial and, in some cases, temporal relationships between anatomical features within and across these imaging modalities" (Computer Graphics, 1987). For example, using the computer a plastic surgeon can modify a patient's features to simulate postoperative results. Such manipulation, based on each patient's diagnostic imagery, can be a powerful tool to help plan a surgery and also allay a patient's anxiety about the outcome. Another emerging application of computer visualization is the custom design of orthopedic reconstructions such as knee replacements through noninvasive 3D imaging.

Such developments in diagnostic imagery dictate a radical departure from conventional methods of teaching and communicating anatomical information. Medicine has traditionally relied on frontal, anterior-posterior views, but this flattened perspective is not sufficient. The explosion of diagnostic imagery has shattered conventions of orientation and requires visualization of oblique, cross-sectional and other unique viewpoints. Using computer-aided design software, students can rotate structures to improve their spatial understanding.

These major changes in spatial representation require heightened attention to fundamental aspects of preparing visuals, such as orienting the viewer. The impression of space can be enhanced by unusual oblique views, but is useful only when the user is properly oriented. Failure to establish the viewer's orientation seriously compromises the communication of the visual, yet we continue to see slides flashed with little or no orienting landmarks or graphic elements. This leaves the viewer with orientation as a first cognitive task rather than proceeding to the intended task of information processing.

Research concerning orientation and mental rotation of figures has provided a body of theory which can potentially be used to solve questions of orientation; however, application of these theories is still sorely lacking. In surgical illustration it is unclear whether it is better to depict a procedure from the surgeon's point of view during the surgery, or whether a view of the patient in anatomical position (upright, anterior-posterior orientation) is best.

Another problem that plagues visual communicators is a lack of standardization of both verbal and visual symbols. Specialty areas often develop representations that are learned by users over time, but comprehensive "dictionaries" of graphic elements would be helpful to assist the new learner and to assure consensus of interpretation.

Standardization of graphic elements would also maximize the amount of information which could be encoded into graphic symbols. For example, illustrators often employ arrows as devices for portraying the idea of direction, movement, or force. What do different types of arrows mean? In medical art there is a tendency to use simple, two-dimensional arrows to imply direction of movement. A three-dimensional arrow can also encode information about force, and can be made more or less monumental to correlate to the amount of force produced. Arrows drawn in perspectives that seem to pierce space can give information about complicated movements such as spirals or rotations. Unfortunately, no standardized vocabulary for graphic elements exists for medical art or for most specialties.

Standardization of data used to construct images would also be a boon to improving accuracy and production efficiency. At present, as each artist begins an illustration he or she must subjectively synthesize information from many resources. A database of morphometric information would assist the artist by providing measurements for an idealized form that can be manipulated, rotated, and embellished using the computer. Following the approach of human factors specialists in the design of tools and environments, the artist would have data sets of measurements to describe the range and standard for forms. Image banks would alleviate the necessity of "reinventing the wheel" (or kidney, brain, or heart in the case of medical art!) every time a new illustration is requisitioned. This way of thinking is somewhat antithetical to the traditional illustrator's mode of thinking, in which the product of artistic labor is considered to be a personal, unique interpretation of the subject matter—a problem that may impede acceptance of stock supplies of imagery.

A project that addresses the issues raised thus far is under way at the Department of Biomedical Visualization at the University of Illinois at Chicago. Aptly named The DaVinci Project, the interdisciplinary research group, consisting of experts from engineering, institutional computing, educational development, supercomputing, urban planning, architecture, medical imaging, and medical illustration, aims to create a RESOURCE CENTER FOR ANATOMICAL IMAGING. Using methods traditionally employed at a microscopic level, the DaVinci Project will establish a comprehensive, accurate description of standard human gross anatomy and its development through time, based on quantitative and qualitative data gathered from diagnostic images and actual specimens (fig. 4). Morphometric analysis and stereology will be used to develop a computer-based stereoanthropomorphic database which can be manipulated, analyzed, and enhanced for various visualization purposes. The database will benefit diverse fields including medical education, bioengineering, anatomical simulator design, forensic science, biological process simulation, surgical instrument design, pharmaceutical research and development, military technology, sports equipment design, and missing persons research.

The DaVinci Project will contribute to teaching efforts, provide a research tool to clinicians and basic scientists, serve as a production tool for artists, integrate diagnostic imagery, and utilize computer technology to standardize and visualize information. Such an endeavor summarily represents the trend toward approaching visual information interdisciplinarily, interactively, and electronically.

# REFERENCES

Doyle, Michael D.; O'Morchoe, Patricia J.; and O'Morchoe, Charles C. C.: A Microcomputer Based Histology Atlas. Proc. Amer. Assoc. Anatomists, 1987.

Evenhouse, Raymond; and McConathy, Deirdre A.: Development and Construction of a Fetal Monitoring Simulator. J. Biocommun., vol. 16, no. 1, Winter, 1989.

Netter, Frank: Ciba Collection of Medical Illustrations. Ciba Pharmaceutical Products, Inc., Colorpress, 1948.

Pernkopf, Eduard: Atlas of Topographical and Applied Human Anatomy, Vols. I, II. W. B. Saunders Co., 1963.

The Urban and Schwarzenberg Collection of Medical Illustrations Since 1896. Urban and Schwarzenberg, 1977.

Visualization in Scientific Computing, Computer Graphics, vol. 21, no. 6, Nov. 1987.
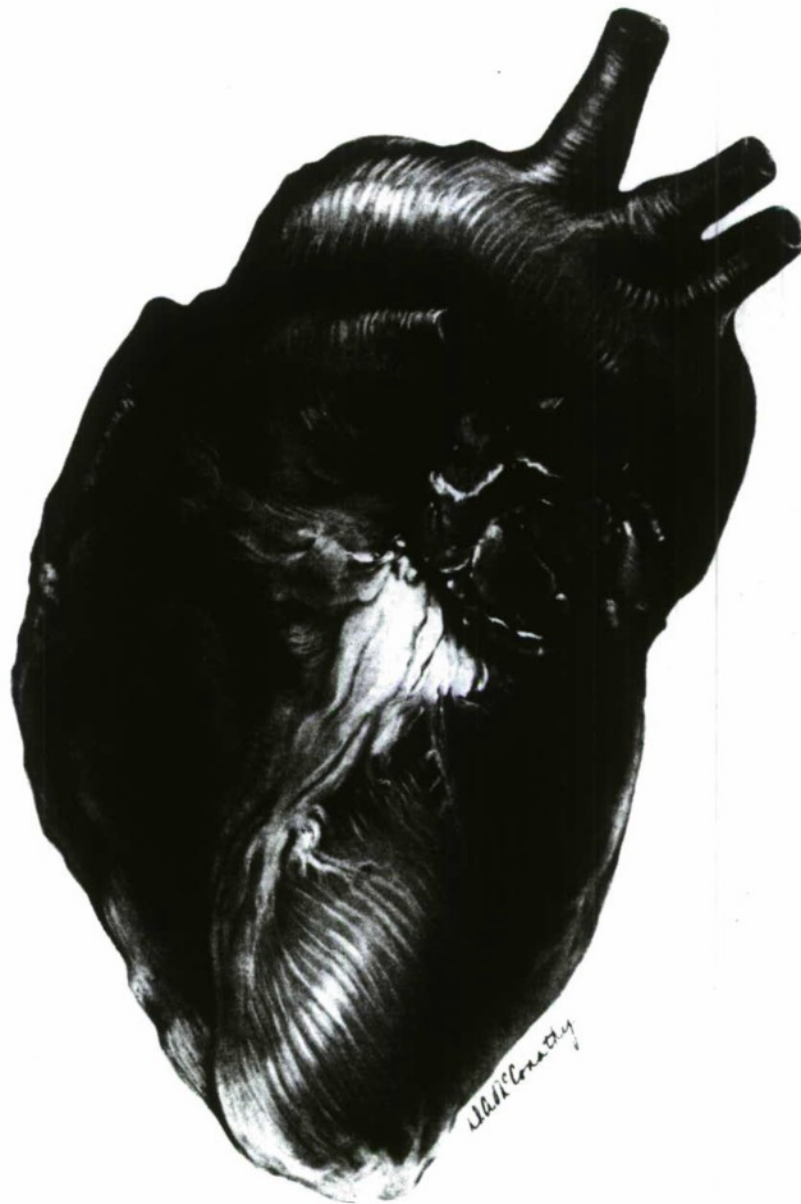
Figure 1.– Traditional medical illustration by Deirdre McConathy, depicting gross morphology of a cadaver heart.
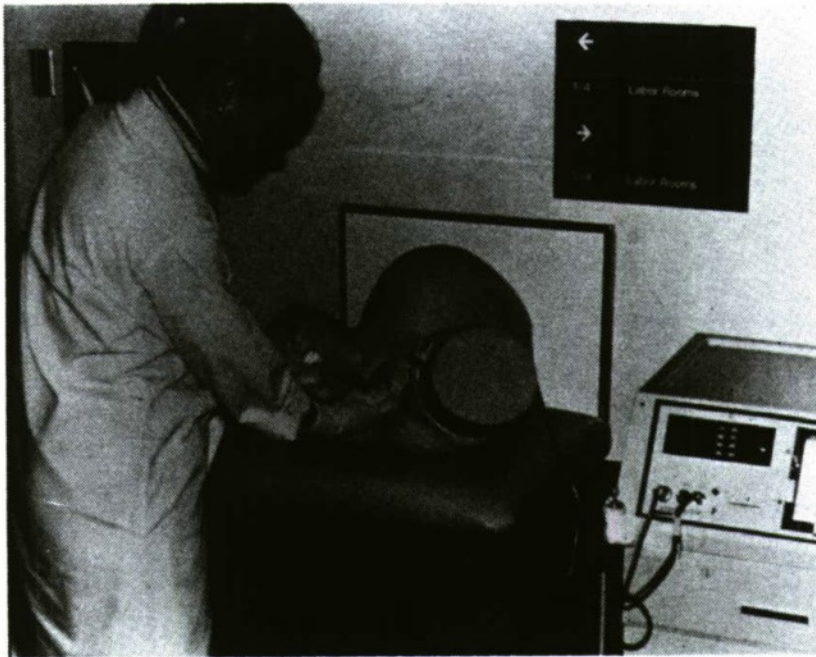
Figure 2.– Interactive patient simulator developed by Evenhouse used to teach instrumentation for fetal monitoring procedure.
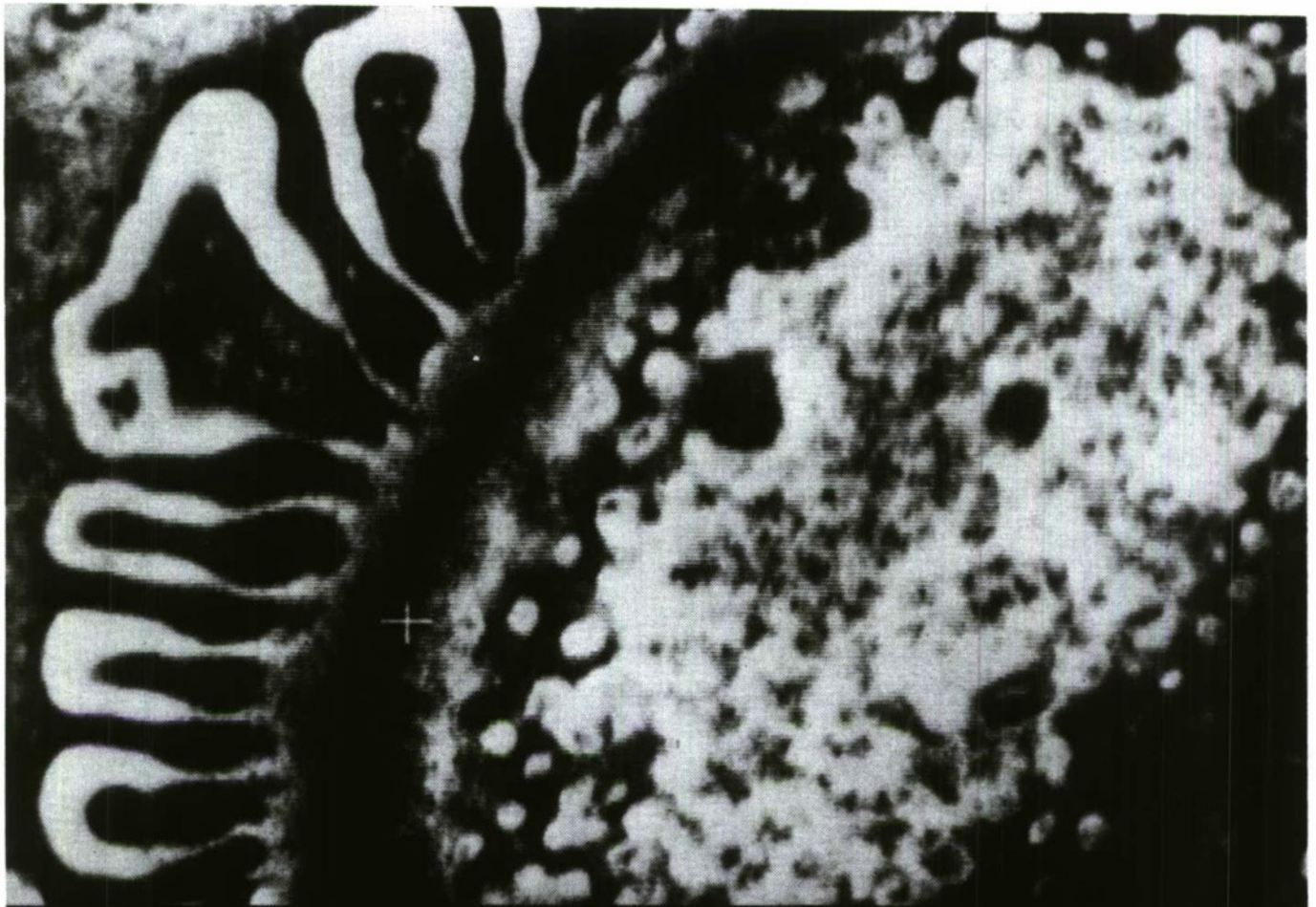
GLOMERULAR BASAL LAMINA
-------------------------------------------------------------------------------

Beneath the pedicels is the basal lamina, about 0.3 micrometers thick.  It
shows a central, electron-dense lamina called the lamina densa.  The lamina
densa is about 0.1 micron thick, with electron lucent layers on external and
internal surfaces termed the lamina rara externa and lamina rara interna.
The lamina densa contains collagen type IV, which acts as a physical filter;
the lamina rara contains glycosaminoglycans rich in heparin sulfate, which
affects passage of of both basic and acidic proteins across the basal lamina.
The basal lamina probably is formed by the podocytes, perhaps with
contributions from the endothelium.

Figure 3.– Electronic textbook developed by Doyle (1987) used to teach medical students about cells and tissues.
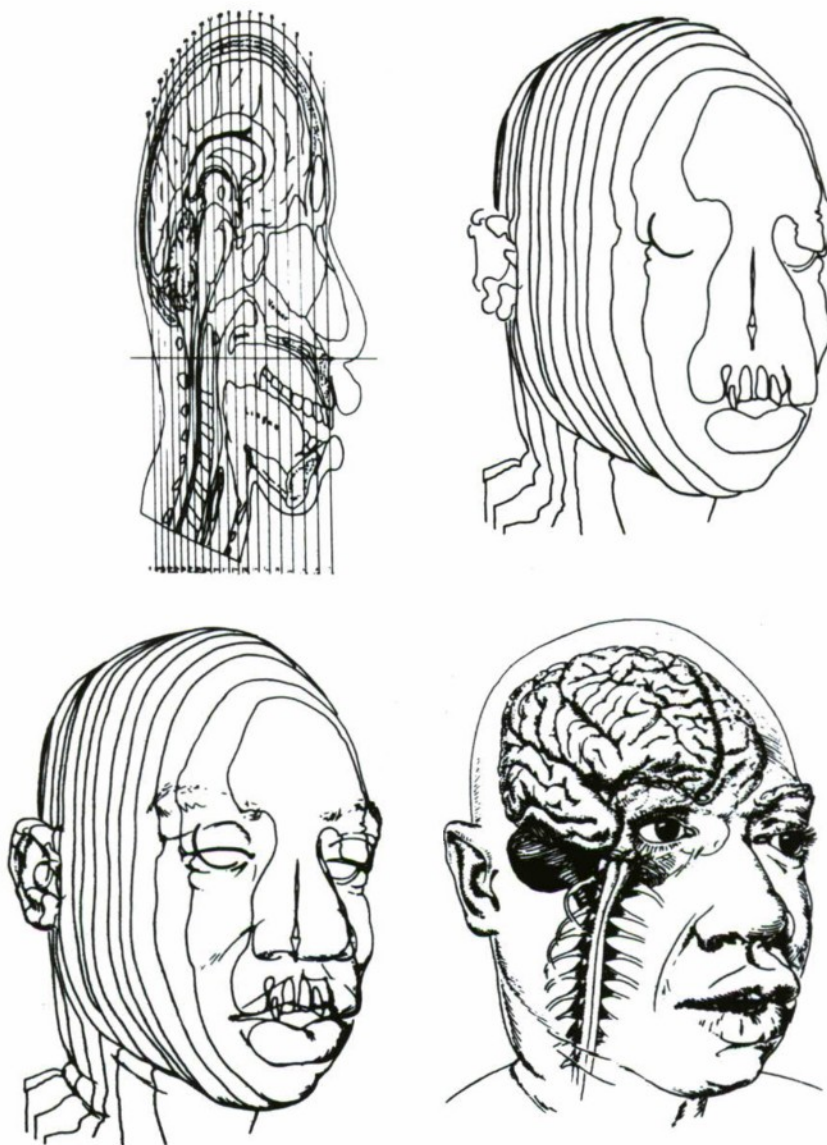
Figure 4.– Representation of the DaVinci Project's aim to use two-dimensional anatomical imaging data to produce a serially reconstructed three-dimensional computer database of standard anatomy.

# THE INTERACTIVE DIGITAL VIDEO INTERFACE

Michael D. Doyle
The University of Illinois College of Medicine at Urbana
Chicago and Urbana, Illinois

A commonly heard complaint in the computer-oriented trade journals is that current hardware technology is progressing so quickly that software developers cannot keep up. As a result, it seems that available applications are always several generations behind in implementing current hardware capabilities. A good example of this phenomenon can be seen in the field of microcomputer graphics.

Today's price/performance ratio is such that an affordable personal computer for a sophisticated user may contain a 32-bit microprocessor operating in the range of 3-4 million instructions per second, an advanced graphics controller capable of 1024x1024 resolution with 256 colors simultaneously displayable from a palette of 16 million possibilities, 2-16 megabytes of RAM and an optical storage device capable of storing 600-800 megabytes of data. Such a system can be purchased today for a price of $10,000 to $15,000. The cost for a similarly configured machine 4 years from now can be expected to drop to the $3000-$4000 range. The physical dimensions of such a machine may shrink from desktop proportions to briefcase size, or smaller.

Such computer systems have the potential for effective storage of, and easy access to, massive amounts of textual and image information. A single optical disk can store all of the text and images contained within a typical set of encyclopedias while providing relatively quick access to any particular information of interest. Optical storage media will most probably supplant many of today's printed forms of publishing.

To effectively exploit the advantages of new mechanisms of information storage and retrieval, new approaches must be made towards incorporating existing programs as well as developing entirely new applications. There exists a great need to integrate more sophisticated graphics into applications and to take a wider view of how that integration can take form.

A particular area of need is the correlation of discrete image elements to textual information. The interactive digital video (IDV) interface embodies a new concept in software design which addresses these needs. The IDV interface is a patented device- and language-independent process for identifying unique image features on a digital video display and which allows a number of different processes to be keyed to that identification. Its specific capabilities include the correlation of discrete image elements to relevant text information and the correlation of these image features to other images as well as to program control mechanisms (fig.1). Very sophisticated interrelationships can be set up between images, text and program control mechanisms using this process.

I originally developed this process during the design of a microcomputer-based interactive atlas of medical histology (histology is the study of microscopic anatomy). Using this system, a medical student can call up from a menu a microscopic image from one of the body's organ or tissue systems. This image is then displayed on the video monitor with no labels or identifying structure names shown. The student can then use a mouse to indicate a particular image area that

he or she would like more information about. Clicking one of the mouse buttons causes the computer to display a screen of explanatory text concerning the particular histological structure indicated (e.g., an individual cell in an image of a group of cells). Pressing the other mouse button would cause the display of a higher-magnification image of that image element (histological structure) selected. The student is then free to interact with this higher-magnification image to obtain further textual explanation or to see even higher magnification views. It should be noted here that this "zooming" capability does not merely involve the higher-magnification display of the same digital image (with the resultant loss of resolution), but rather causes the display of an entirely different image with no decay in resolution or image quality. For example, if the on-screen image was of a 1000X light microscopic view of some tissue, selecting the "zoom" feature would cause the display of a low-magnification (3500X) electron microscopic image of that particular type of structure. These correlations can be caused to run in reverse, so that the student could zoom from high magnifications to lower magnification views or he or she could enter the name of a structure from the keyboard with the resultant display of an image containing the highlighted structure on the video display.

Image databases adapted for the IDV interface are extremely memory-efficient. The data storage load for a single image and correlation mechanism is less than 1% larger than the original compressed image file before adaptation to the system. It is therefore practical to include all of the 1200 or so images needed for a complete histology atlas on a single CD-Rom disk. Another advantage to the process is that it runs very quickly, and this speed is not affected by the resolution of the image. The histology atlas runs very fast on an unadorned IBM PC (4.77 Mhz) with the appropriate graphics controller and disk storage device. Although the images in the atlas are only 512x484 pixels in resolution, the program would achieve the feature identification just as quickly if the image resolution were 4000x4000 pixels.

A specific objective in the development of the Interactive Atlas of Histology was to eliminate the distraction of having all of the important discrete elements within an image labeled on the screen and yet maintain the capability for immediate access to the exact descriptive textual information which the student desires.

In some situations, computer graphic images can contain so much information that it is not practical or not necessary to see all of the text-based information relevant to a particular image. Such a case exists in the graphic display of supercomputer-level image output. The IDV interface could be of great practical value in allowing the scientist to interact directly with the graphic display of, for example, a complex biological process simulation. A custom-designed interface could allow the researcher direct and immediate control over program flow for a simulation while it is executing, or immediate textual elaboration on an interesting feature of the simulation output display.

Head-up displays are currently of great interest in the aerospace industry. These displays have the effect of placing the user within the virtual environment of the computer image. A great deal of research is being done towards making the user interface for such a display as intuitive as possible. Techniques such as retinal scanning are being investigated as possible means to achieve a very natural-feeling way to specify a location within the display. The IDV interface would be an effective way to correlate this intuitive locator mechanism to desired relevant computer responses.

Other possible applications for the process are numerous: computer-aided education for information-intensive fields such as medicine or the military, for the earliest educational levels or

for remedial or special education; image-based reference works such as atlases, catalogs, maps or navigation systems; cognitive rehabilitation systems, for head injury or Alzheimer's patients, to build associative relationships and still allow a controllable degree of freedom of interaction; interactive art displays; foreign language education systems; and entertainment programs or games.

The IDV interface is an attempt to redefine the role that computer graphic display images play in the function and purpose of application programs. It extends the concept of interaction to allow a user to interface directly with an image and not be distracted by unwanted information or the mechanics of computer operation. Although my own interests in developing applications with this process are limited to educational computing, it is my hope that others will undertake to explore its integration into the myriad of possible interactive graphic applications for which it is so aptly suited.

Relevant Textual
Information

INTERFOLLICULAR LYMPHATIC ENDOTHELIUM

The interfollicular lymphatic capillaries have a
thin walled endothelial lining. This lining
displays a discontinuous basal lamina, anchoring
filaments and a lack of fenestrations when viewed
at the ultrastructural level.
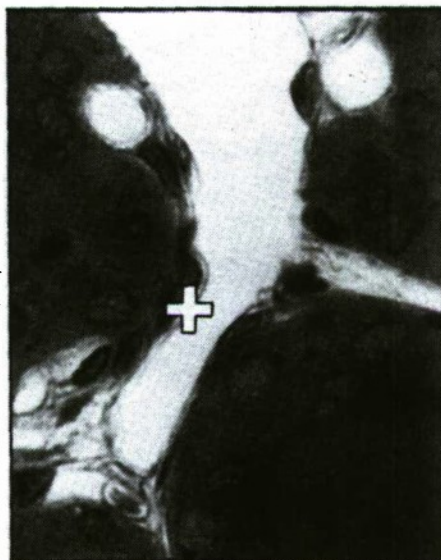
Other Images

Original Image
Element
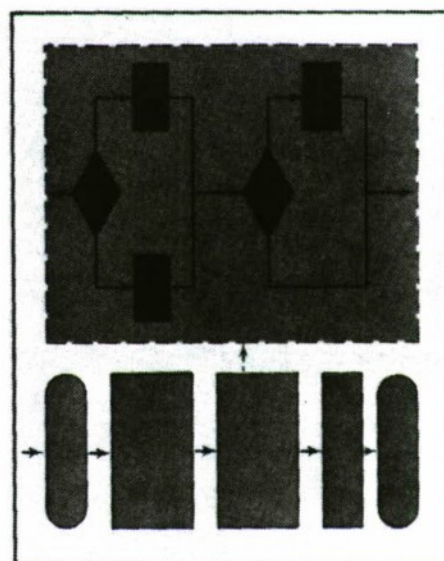
Program-Flow
Control Mechanisms
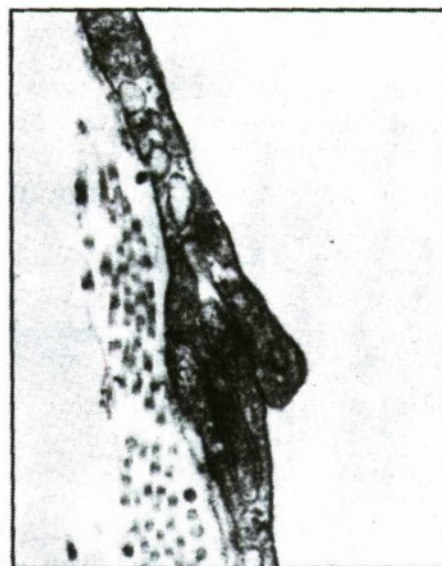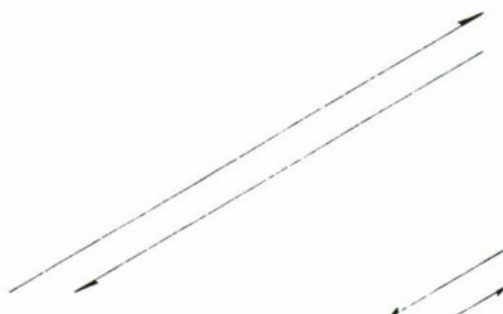
Figure 1.- The IDV interface allows very sophisticated interrelationships to be set up between
images, text and program-control mechanisms.

Pending publication in *Perception and Psychophysics*,
the paper "Efficiency of Graphical Perception"
by Gordon E. Legge, Yuanchao Gu, and Andrew Luebker
has been withdrawn from this *Proceedings*[1]

---

[1] If copyright permission is obtained before printing, we will add this paper as an addendum.

# DISPLAYS, INSTRUMENTS, AND THE MULTI-DIMENSIONAL WORLD OF CARTOGRAPHY

George F. McCleary, Jr.
Department of Geography
University of Kansas
Lawrence, Kansas

Cartographers are creators and purveyors of maps. Maps are representations of space—geographical images of the environment. Maps organize spatial information for convenience, particularly for use in performing tasks which involve the environment. There are many different kinds of maps, and there are as many different uses of maps as there are spatial problems to be solved.

## MAPS AND THE DISPLAY-INTRUMENT DICHOTOMY

The many different uses of maps can be categorized into two groups. Some maps are used passively—they *display* information. They are subjected in some cases to only a glance, a moment of study, and little more; in some situations (although the author would no doubt prefer otherwise) they seem to be ignored. Information obtained from maps used as displays is gained by visualization—the eye-brain system processes the display without assistance from any device (e.g., ruler, planimeter).

Other maps, in order to fulfill their missions, must be studied, analyzed or measured. They are used as *instruments*. This is clearly the case with maps used in sea or air navigation or those used to carry out engineering operations. Map use in situations like these is an active process and the map cannot be ignored—it is used with precision, and the efficiency of performance of the task in which it is used depends, sometimes entirely, on the accurate use of the map.

The two parts of figure 1 indicate these extremes: A simple location map from a newspaper contrasts in many ways with the level of detail and the utility of the navigation chart (here shown not only with water depths and graticule marks, but also with electronic navigation system information). While these illustrations make this dichotomy obvious, this difference in approach to examining the uses of maps and to the understanding of the cartographic process presents a significant opportunity for clarifying concepts and procedures which have tended to be passed over by cartographers.

The approach taken here to the display-instrument dichotomy is not contradictory to that set forth by Ellis (1987), but it departs from his perspective in two ways. First, it is applied *only* to maps. Second, the focus is on the *use* of maps—not their creation.

Ellis considers all maps to be instruments, but there are some maps which must clearly be *displays,* even from his perspective. The very large paintings by Jasper Johns come immediately to mind (Crichton, 1977), along with those maps used quite often as a major element of the message in either advertisements or portraits—in these cases, the map serves a simple (often propagandistic) role, for it lends worldly credibility to the person or situation involved.

The representation of the land surface provides an excellent illustration of the display-instrument dichotomy in map creation and use. To create any map a considerable amount of data is required; for a long time there were no significant data available to create a detailed map of the land surface. At the outset there was only the relative location of the feature and some characterization of it (e.g., "over there, a hill"). At this point, there was no real need for a more detailed description. As science and technology developed, it became possible and necessary, first, to locate things more precisely (the graticule and other coordinate reference systems, as well as horizontal and vertical datums were established) and, second, to describe the surface of the land in more systematic terms (verbal characterizations yielded to graphic symbols in a map format, then to representations of slope and finally, with the availability of data, to the mapping of elevation using contours) (Hodgkiss, 1981; Harvey, 1980). The sequence of illustrations in figure 2 provides some high points in this evolution.

The inventory of techniques presented here ends not with the contour—an *instrument* for land surface representation—but with a shaded relief map. While the industrial revolution and the emergence of industries which required large quantities of natural resources needed the kind of information about the land surface that only contours could provide, another aspect of the land surface rose to importance. The contour provides a representation of the land surface suitable for measurement—it is an instrument, and it is a very poor device for *visualization*—it does not create a good display. It is difficult, even impossible, for even a sophisticated map reader to gain a good overall image of the landscape from a topographic map. Therefore, in a number of different map use situations where visualization of the characteristics of the land surface is important, cartographers have employed shaded relief methods on their maps.

The problems associated with land surface representation illustrate nicely the interrelationships among a culture, its science and technology, and the maps which were developed. Different cultures and different times generate different needs for maps, and cartographers have responded to these needs in different ways.

Consider the problem of accomplishing a single task—accurate sea navigation. What form of map—instrument—was and is available? At the outset there were probably no maps (as we understand the concept of the map as a two-dimensional representation); there were only verbal (at first oral and then, later, written) instructions. These yielded to the portolan charts, which codified the relationship between the magnetic "environment" and the land-seascape (given an origin and a destination, there is a straight-line magnetic course between them). While determining latitude has been understood for several thousand years, celestial navigation requires an accurate measurement of time to determine longitude—and the whole process required two instruments: the chronometer and the cylindrical conformal projection. While the former is an eighteenth century invention (Harrison won the prize awarded for creating the first accurate nautical timepiece, and LeRoy and Earnshaw made major innovations which made the chronometer more reliable and inexpensive; see Brown, 1949, and Bowditch, 1966), the latter was first used by Mercator in 1569 (and mathematically described by Wright in 1599; an earlier use, by Etzlaub in 1511, is much less notorious than that by Mercator (Maling, 1973)). Other navigation instruments came much later, including, for example, the electronic navigation LORAN system, and the inertial guidance and satellite-based systems in use today (Monmonier, 1985).

This sequence of development is presented in figure 3: Descriptive guide, portolan chart, Mercator projection, LORAN network, and so on. The final element in this sequence of instruments is a display: a map from an advertisement for a cruise. The sequence of development in

navigation, piloting, dead reckoning, celestial navigation, electronic navigation, and—for the tourist—vicarious navigation, is mirrored by a sequence of instruments (and one display).

## CATEGORIES OF MAP USE

While there are many different classification systems which have been created for maps, none take advantage of the display-instrument dichotomy. In terms of map use, this dichotomy can be paired with another to create a four-category system of map use. Maps are used either for navigation or for environmental management. One either uses a map to go from one place to another, or the map is employed to provide information about the environment, either for the sake of the information itself ("this map shows the major battles in the European theater in World War II") or so that the information can be used to organize or modify the environment (a map of election precincts or a house plan).

In most cases the navigation map is an instrument. In advertisements, travel guides, and the like, however, it is used as a display. An increasing number of maps are being produced as displays for environmental management; these occur not only in the news media, but also in professional and educational journals and books. Few, if any, of these require the analytical and measurement capabilities of the engineer's plan or the architect's drawing. As displays, these maps require the properties necessary for effective visualization. In such a case, the focus of the map creation process shifts from processes which are founded principally on geometric and geographic precision to those which accommodate the human eye-brain (visual information processing) system.

These four map use categories are compared in figure 4.

## THE CHARACTERISTICS OF MAPS

Maps have many characteristics, but all fall into two categories: They are either aspects of the *structure of the map*—those things associated with the scale of the map and its "projection," or they are related to the *content of the map*—the graphic symbols which represent the features of the environment portrayed.

### Structure: Space and its Transformations

The literature on map projections is extensive; here we find problems that have confounded and captivated the minds of cartographers for centuries. One will find in any single source only a few of "the answers," for as the uses of maps are very different, so too are the projections which have been used (and misused) for these different requirements. Some fundamental concepts will, however, enable us to resolve the projection problems in terms of the display-instrument dichotomy.

The focus of the cartographic interest in projections has been on the transformation of the spherical earth to the plane. Here, after reduction to some particular scale, the primary

considerations are the *properties* of the different transformations. For navigation at the instrument level, the Mercator projection comes immediately to mind. There are, of course, other projections used for navigation, and most of these are, like the Mercator, conformal; i.e., all angles are represented correctly. The Mercator projection is unique, however, for it is only on this projection that all rhumb lines (lines of constant compass direction) are shown as straight lines—an extraordinarily useful situation for a navigator.

There are, however, a number of other facets of the Mercator projection which make it very important to this discussion. First, it does not show great circles as straight lines (this is the property of the gnomonic projection—the gnomonic is the traditional companion to the Mercator; on it all straight lines are great circles—one plots the great circle route between two points, then compiles this path on the Mercator as a set of rhumb lines which are used in the navigation process.) Second, in the transformation of the spherical surface which is required to develop the property of conformality, the Mercator projection exaggerates the sizes of areas; this is a problem which has caused great difficulty when this projection has been used for maps of the world designed to display statistical data. It is a problem which has existed for several hundred years; like the durability of Greek scientific concepts in the Renaissance, it is the Mercator image of the world which has become the consensual view of people around the world. What General Frederick Morgan recognized as a key problem in gaining American support for Operation OVERLORD (Morgan, 1950) (fig. 5) has been documented in depressing detail by Saarinen (1987) (fig. 6).

The solution to the display problem is simple: If you are to make a map of the surface of the Earth, a display to provide information for visualization about some aspect of our environment, use an equivalent (equal area) projection. Here areas on the surface are shown in correct proportion. This has been done—and done again—and again. Unlike the Mercator, the cylindrical conformal projection, there is no unique solution for the cylindrical equivalent projection—there are a variety of possibilities. Further, when one relaxes a constraint on the transformation process, then an even wider array of possibilities emerges. While many have "solved" the problem once, others have created a series of solutions, all unique and all useful. None of these has, however, achieved universal acceptance. Why? None of them looks enough like the Mercator—the consensual—image of the world.

There are many equal area projections (fig. 7), and there are a growing number of compromises: projections which are neither conformal, not the Mercator, nor equivalent—just something between these two, with none of the properties of either. The compromise by Miller is widely used (Snyder, 1982)—it is not equal area, but it has a lot of Mercator-like properties. The one developed by Robinson (1974), and termed "orthophanic" (it "looks correct"), is based on several decades of study of the problem, and the author recognized (and published) its limitations. This is in marked contrast to the campaign mounted by Peters (1983) in support of his equal area projection—the list of "fidelities" associated with it are an insult to those who understand, but a great lure to those who seek a single solution to a problem which has none.

The final event in the organization of the structure of maps is the work with "cartograms"—topological transformations of geographic space on the basis of some set of statistical data. The sizes of areas (countries, states, etc.) are functions of their populations, economic level, or some other statistical measure (Tobler, 1963). Cartograms of this type are a recent invention (Raisz, 1934), but their navigational counterparts date to the Crusades. Automobile strip maps, the distorted maps used by railroads (and many rapid transit systems), and the diagrammatic maps employed by airlines are not only useful, but they are often much easier to understand (be it to

visualize or to measure) than their geographically correct counterparts. They represent, as well, a sophistication in the handling of map structure well beyond the normal transformations (projections) generally employed. The earliest cartograms—maps based on a structure of a conceptual space—are the T-in-O maps. These medieval *mappae mundi*, generally considered as perpetrators of myth and dogma, simply reflect a view of the world organized more on the basis of theology than geography (Wilford, 1981) (fig. 8).

In handling the structure of a map (as either a maker or a user), one must turn to fundamentals in order to make an appropriate decision. Choose first the projection which has the properties necessary for the use of the map (conformal for navigation and surveying, equivalent for visualization of statistical information, or one of many other properties—such as equidistance—if the use requires it). Given the important property, then select the least distorted version possible (Robinson et al., 1984).

## Content: Data and Their Transformations

Spatial—environmental—information can be conveyed in a number of different ways. One can use words, either written or spoken. Numerical data can be employed, and one is often confronted with great quantities of tabular data, all organized in a form more appropriate for an accountant than for an environmental analyst. These forms, among others, are found in the categories of what Moellering (1980) has called "virtual maps." In some cases verbal or numerical environmental descriptions—maps—are more effective for handling a task than "a real map"—a graphic description. In most situations, however, maps are much more effective for representing the environment, either for display or for use in measurement.

The question which concerns many people, however, is just how effective are these graphic displays. Are they understood more accurately than the verbal essay or the statistical table? While there is a legacy of nearly two centuries of "thematic maps" (Robinson, 1982), it has only been in the last half century that serious consideration has been given to the problems associated with reading—visualizing—these maps. It was only in 1967 that Jacques Bertin described and explored the six visual variables, the graphic vocabulary (Bertin's work was made available in English in 1983). While it is possible in 1988 to present information using graphic devices that provide a reasonable expectation that the message will be communicated appropriately, it is clear that other forms of presentation will fail to achieve the goal.

The six visual variables are illustrated in the ways that they can be used to represent point, line, and area data in figure 9.

It is not possible here to analyze the entire situation, but the use of symbol size (graduated circles) is illustrated in figure 10. In the first map, the sizes of the circles are directly proportional to the populations of the Kansas and Missouri counties which they represent; a circle representing 10,000 people is twice the size of a circle representing 5,000 people, and a tenth the size of one representing 100,000 people.

A large number of studies have shown that the human eye-brain (visualization) system does not respond to these circles in the same way that a mathematical measuring device would; it is clear that circle size differences are underestimated (Stevens, 1975). The second map compensates for this characteristic of the human system; the size of the smallest circle is the same here as on the first

map, but all other sizes have been rescaled to overcome the size difference underestimation. Note that the largest circles are significantly larger here than on the first map; the map has been developed with the human eye-brain system as the focus—the numerical data have been transformed to a visual series which should present the information correctly to most map readers (McCleary, 1983).

This is a short, and highly simplified, explanation of a very complex problem. To do justice to it, one needs to explore each visual variable, alone and in combination and context. Each added factor makes the visualization situation more complex. In the same way that the addition of an adjective as a modifier to a noun changes the understanding of the noun (and the addition of an adverb modifies the idea even further), the use of visual variables in combination changes the message to the map user. When a symbol is placed in a context, it—like the noun phrase placed in a sentence or a paragraph—may assume a different meaning. There is a great amount of research to be done before there will be a clear understanding of all the processes and responses to problems in the visualization of maps. Achieving an understanding of the graphic vocabulary and adapting this knowledge to the many variations in graphic displays should not, however, dissuade people from developing and using innovative methods for information. Whether it be for a display or for an instrument, some new approach might elicit more appropriate user behavior for a particular task than a device or procedure which has a legacy of extensive use.

If one learns to write better by reading extensively, one will for certain be better prepared to present data on maps if he or she "reads" widely, examining maps in many different places, in many different forms, for many different purposes.

To this end, the reader might explore the work presented in several volumes. The statistical textbook by Schmid and Schmid (1979) provides a traditional benchmark approach. From the cartographic perspective, Dickinson (1973) focuses directly on the merger of statistics and maps. Monkhouse and Wilkinson (1971), on the other hand, provide an in-depth exploration of mapping techniques. The encyclopedic approach here contrasts greatly with the technical approach used in nearly all of the other cartographic textbooks available; see, for example, *Elements of Cartography* (Robinson et al., 1984).

Lockwood (1969) ranges among a wide variety of maps and graphs, while Fisher (1983) focuses on fundamental facets of the mapping problem. Herdeg (1982) has collected a wide array of material from an even wider array of resources. Southworth and Southworth (1982) focus on maps—a "scrapbook" approach. One might accompany their exploration of these with the text on *Map Appreciation,* by Monmonier and Schnell (1988); this volume focuses on types of maps. *Map Use,* by Muehrcke (1986), is more concerned with process.

All of these volumes have much to recommend them; all have their liabilities. Cartography is a field in transition. Maps are not the property of the product of the cartographer alone. In fact, as some of these volumes indicate clearly, innovation (and the associated excitement) occurs quite often outside the realm of the professional mapmaking clan.

## THE HUMAN, MAPS, AND BEHAVIOR

All of the discussion which has gone before has ignored a major area of activity in carto-graphic research and instruction: cognitive mapping. Here, and in the other research areas associ-ated with it (including environmental psychology, environmental cognition, and the like), the attention lies clearly on the maps which are integral components of the human system. Those who study cognitive maps are concerned with the characteristics of the maps "housed" in the mind of an individual, with the origins of these maps, including different sources of information and the envi-ronment, as well as with the behavior which is associated with the uses of these mental images (Downs and Stea, 1977).

This can be explained very simply in a diagram. Humans interact with the environment; on the basis of this interaction, information is transmitted from the environment. This information results from direct interaction with the environment as well as from resources (of all types) which describe the environment. This information can be said, simplistically, to form the basis for a *cognitive atlas,* a collection of maps resident in the mind of the person. While the contents of the atlas are derived principally from the environment, either directly or vicariously, the human imagi-nation is often used in the same way that cartographers have always imaginatively filled the blank spaces on maps (fig. 11).

The "bottom line" in this process is the human response to the environment, the behavior which results from the application of a cognitive map in the solution of some environmental prob-lem (McCleary, 1987). When map use is direct, and very significant to some environmental prob-lem, the map will no doubt have a major effect on the behavior. (This has been demonstrated in a number of ways, in problems of different types; see McCleary and Westbrook (1974) for a very direct analysis of this system.) In many instances, however, the role of the map may be less obvi-ous; as we have seen throughout this discussion, the impact of a map may be reflected in many subtle ways.

## CONCLUSION

The world of the cartographer is one of many dimensions and complications. There are not only problems in understanding map structure (projections) and content (symbols, as well as the design of the map), but there is also a continuing series of changes in needs and requirements. Accompanying all of this there is the ever-present change in technology—and an evolving philoso-phy for the discipline.

What is significant here is that Ellis has provided one more way to "tie down" various parts of the map problem: some maps are displays, while others are instruments. This has been true from the beginning, but a clear recognition of these two major components of the cartographer's dichotomous existence and an implementation of this view in our teaching, research, and produc-tion—as well as in the philosophizing—should help a great deal in organizing the enterprise.

# REFERENCES

Bertin, Jacques (trans. by William J. Berg), Semiology of Graphics: Diagrams, Networks, Maps, 1983. Madison: Univ. Wisconsin Press.

Bowditch, Nathaniel, American Practical Navigator: An Epitome of Navigation, 1966. Hydrographic Office Publication No. 9; Washington, GPO.

Brown, Lloyd A., The Story of Maps, 1949. New York: Bonanza Books.

Crichton, Michael, Jasper Johns, 1977. New York: Harry N. Abrams.

Dickinson, G. C., Statistical Mapping and the Presentation of Statistics, 1973. Second ed.; London: Edward Arnold.

Downs, Roger M. and Stea, David, Maps in Minds: Reflections on Cognitive Mapping, 1977. New York: Harper and Row.

Ellis, Stephen R., "Pictorial Communication: Pictures and the Synthetic Universe," 1987. Presentation to Spatial Displays and Spatial Instruments Conference; Asilomar: NASA Ames Research Center and U.C. Berkeley.

Fisher, Howard T., Mapping Information: The Graphic Display of Quantitative Information, 1982. Cambridge, MA: Abt Books.

Harvey, P. D. A., The History of Topographical Maps: Symbols, Pictures and Surveys, 1980. London: Thames and Hudson.

Herdeg, Walter, ed., Graphis Diagrams: The Graphic Visualization of Abstract Data, 1981. Zurich: Graphis Press.

Hodgkiss, Alan, Understanding Maps: A Systematic History of Their Use and Development, 1981. Folkstone: Dawson.

Lockwood, Arthur, Diagrams: A Visual Survey of Graphs, Maps, Charts and Diagrams for the Graphic Designer, 1969. London: Methuen and Co.

Maling, D. H., Coordinate Systems and Map Projections, 1973. London.

McCleary, George F., Jr., "An Effective Graphic 'Vocabulary'," 1983. IEEE Computer Graphics and Applications, 3 (2), 46-53.

McCleary, George F., Jr., "Discovering Cartography as a Behavioral Science," 1987. J. Environment. Psychol., 7, 347-355.

McCleary, George F., Jr. and Westbrook, Nicholas, "Recreational and Re-Creational Mapping," 1974. Worcester, MA: Clark Univ. Cartographic Laboratory.

Moellering, H., "Strategies of Real-time Cartography," 1980. Cartographic J., 17, 12-15.

Monkhouse, F. J. and Wilkinson, H. R., Maps and Diagrams, 1971. Third ed., Methuen and Co., London.

Monmonier, Mark Stephen, Technological Transition in Cartography, 1985. Madison: Univ. Wisconsin Press.

Monmonier, Mark and Schnell, George A., Map Appreciation, 1988. Englewood Cliffs: Prentice-Hall.

Morgan, Sir Frederick, Overture to OVERLORD, 1950. New York.

Muehrcke, Phillip, Map Use: Reading, Analysis, and Interpretation, Second ed., Madison, WI: JP Publications, 1986.

Peters, Arno, The New Cartography, 1983. New York: Friendship Press.

Raisz, Erwin, "The Rectangular Statistical Cartrogram," 1934. Geogr. Rev., 24, 292-296.

Robinson, Arthur H., "A New Map Projection: Its Development and Characteristics," 1974. Intern. Yearbook Cartogr., 14, 145-155.

Robinson, Arthur H., Early Thematic Mapping in the History of Cartography, 1982. Chicago: Univ. Chicago Press.

Robinson, Arthur, Sale, Randall, Morrison, Joel and Muehrcke, Phillip, Elements of Cartography, Fifth ed., 1984. New York: John Wiley and Sons.

Saarinen, Thomas, "Students' Maps Reflect Geographical Bias," 1987. Washington: National Geographic Society.

Schmid, Calvin F. and Schmid, Stanton E., Handbook of Graphic Presentation, Second ed., 1979. New York: John Wiley and Sons.

Snyder, John P., Map Projections Used by the U.S. Geological Survey, 1982. USGS Bull. 1532; Washington: GPO.

Southworth, Michael, and Southworth, Susan, Maps: A Visual Survey and Design Guide, 1982. New York Graphic Society Book; New York: Little-Brown.

Stevens, S. S., Psychophysics: Introduction to its Perceptual, Neural and Social Prospects, 1975. New York: John Wiley and Sons.

Tobler, Waldo R., "Geographic Area and Map Projections," 1963. Geogr. Rev., 53, 59-78.

Wilford, John Noble, The Map Makers, 1981. New York: Alfred A. Knopf.

**DISPLAY**

**INSTRUMENT**
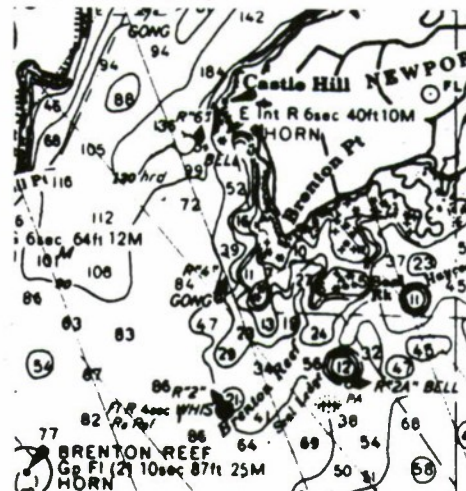
Figure 1.– The map as a display.  Left:  A newspaper map (<u>Christian Science Monitor</u>).  The map as an instrument.  Right:  A coastal chart (National Ocean Survey).
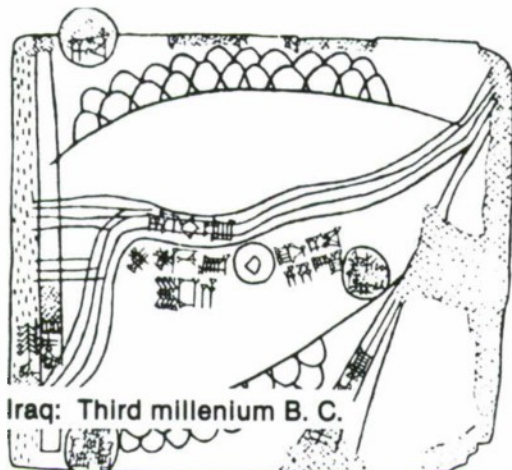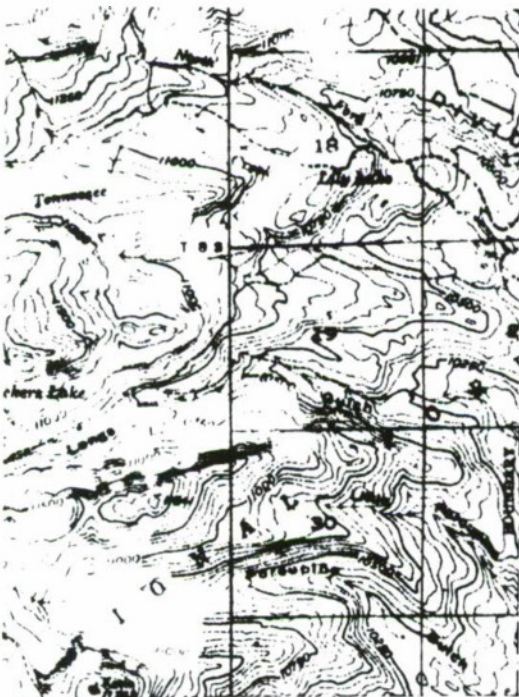
Figure 2.– Evolution of mapping the land surface. Upper left: Ancient map from a clay tablet (outline sketch, with mountains shown in horizontal perspective), with portion of "Sabaundia et Burgundiae" from Abraham Ortelius, <u>Theatrum Orbis Terrarum</u> (a simplified oblique view of hills and mountains). Upper right: Portion of a Swiss topographic quadrangle, using hachures to indicate slope. Lower left: From the U.S. Geological Survey, contours used to represent elevation—and (lower right) a shaded relief version of the same map.

From a pilot's guide, 1691

e de la Riviere de Sei côtoye étant au Noi
Nord Nord Oüeſt d elle 3 ou 4 lieües.

de la Cosa, 1500

Mercator (outline sketch), 1569

Departing Acapulco 3:23.
Departing San Juan 4:2.

St. Thomas

Juan

Antigua

Cartagena

Aruba

Princess Cruises, 1987

United States (sailing chart), 1970

SAN FRANCISCO
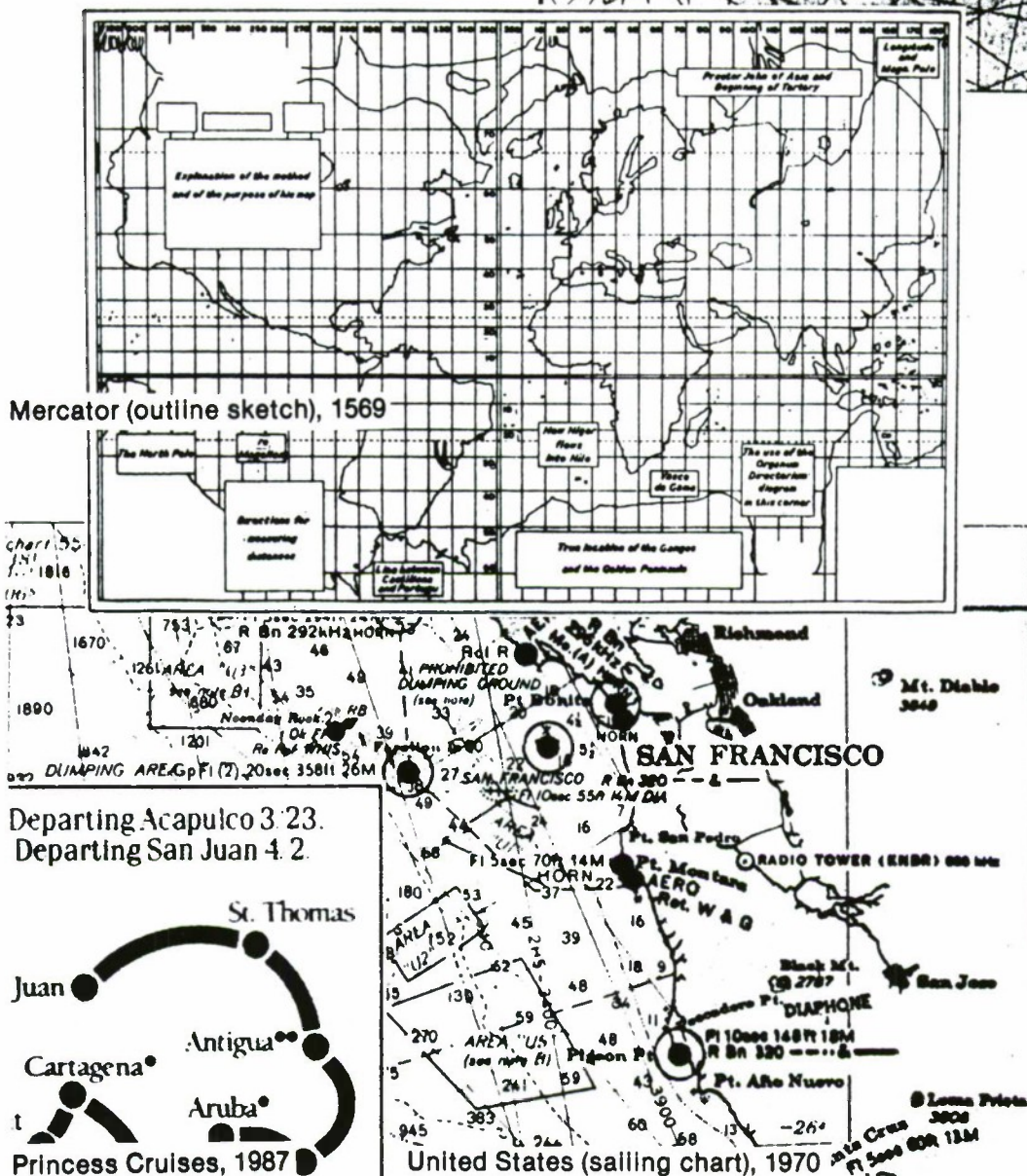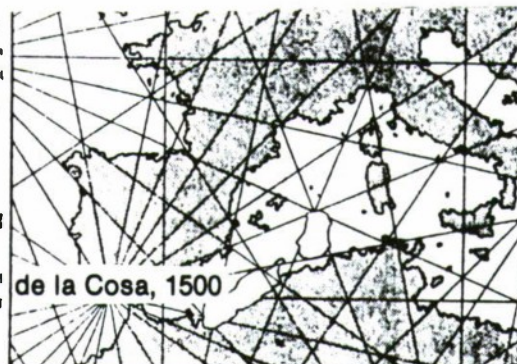
Figure 3.– Sequence of examples showing the evolution of maps used as instruments for navigation. Upper left: Portion of a pilot's guide. Upper right: An outline sketch of a portion of Juan de la Cosa's portolan chart. Center: An outline sketch of the Mercator world map. Lower: Portion of a sailing chart (from the U.S. National Atlas, 1970)—with a map from an advertisement for a Caribbean cruise.

**DISPLAY**

HOGS
NUMBER. 1959

**INSTRUMENT**

**ENVIRONMENTAL MANAGEMENT**

1 DOT=10 000 HEAD
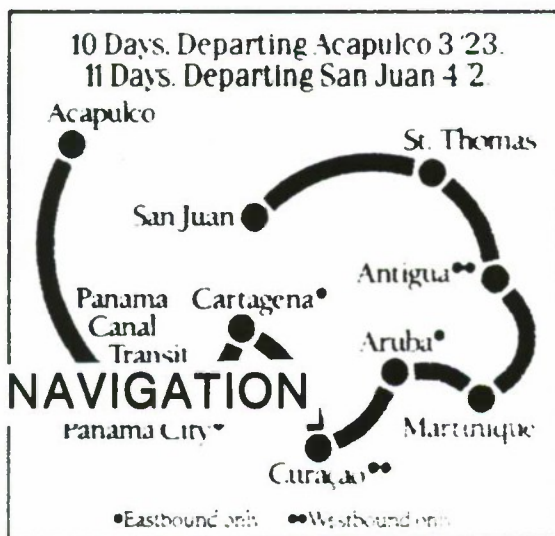
**NAVIGATION**

Figure 4.– The four categories of map use. Upper left: Environmental management, display (U.S. Depart. of Agriculture). Upper right: Environmental management, instrument (portion of an engineering drawing, Army Corps of Engineers). Lower left: Navigation, display (from an advertisement by Princess Cruises). Lower right: Navigation, instrument (portion of the "upside-down map" from New York to Florida produced for the ESSO Company).
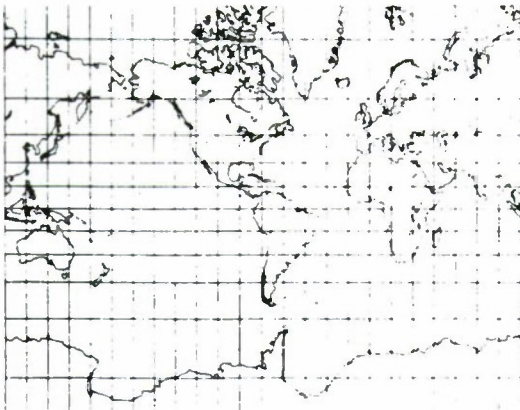
CYLINDRICAL (Mercator), 1569

MERCATOR
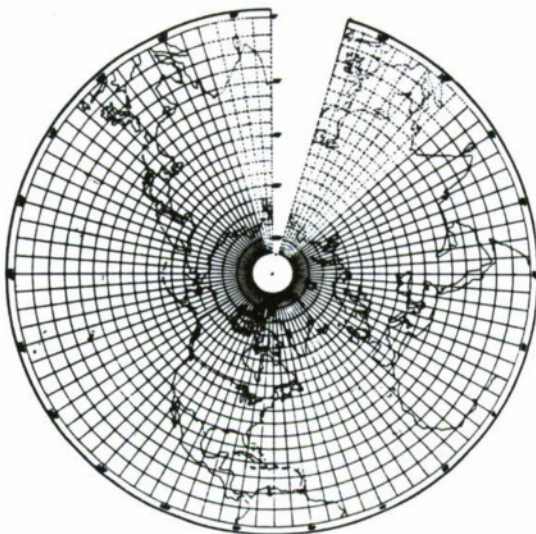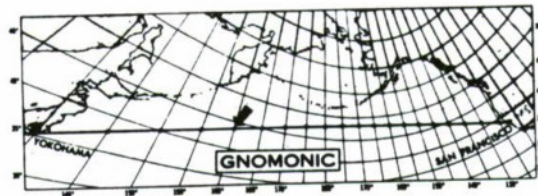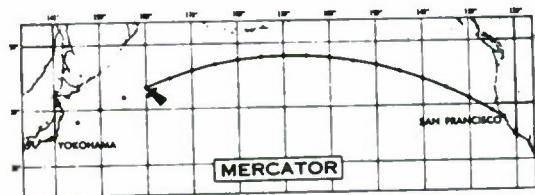
GNOMONIC

CONIC (Lambert), 1772

GNOMONIC (Known before 600 B. C.)

PLANAR (STEREOGRAPHIC, Hipparchus), 160-125 B. C.

Figure 5.– Conformal projections and the gnomonic: instruments used for navigation.

## Students' Maps Reflect Geographical Bias

Sketched from memory, these world maps are a sampling of the 4,277 drawn by first-year college students in 54 countries. Geographer Thomas F. Saarinen collected the maps to test mental images of the world – many not too accurate. Most students centered Europe; some, Asia or the Americas. Nearly all took the task seriously, but a few saw it as a chance for humor. The mapping project was sponsored by the International Geographical Union and financed by the National Geographic Society.
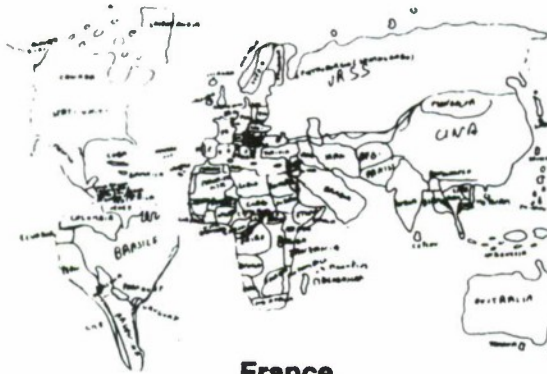
**France**

**Indonesia**

**United States**

**Australia**

**Japan**

**Philippines**

©1987 NATIONAL GEOGRAPHIC SOCIETY

Figure 6.– A summary of key points from a research study by Thomas F. Saarinen:  Mental images of the world are generally organized very similarly, no matter where the student lives—the basic organization is a sixteenth-century perspective:  The Mercator structure of the world.

26-15

EQUIVALENT PROJECTIONS
AND COMPARISONS

Sanson-Flamsteed (Sinusoidal), 1606

Lambert, 1772

Mercator (conformal), 1569

Mollweide, 1805

Hammer-Altoff, 1892

Miller, 1942

Eckert IV, 1906
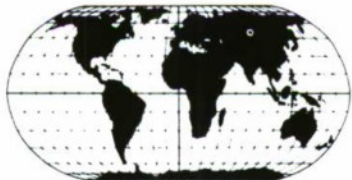
Robinson (orthophanic), 1963

Behrmann, 1910

Boggs (Eumorphic), 1929

Peters, 1967

Figure 7.– Seven equivalent projections, from 1606 to 1929. For comparison, note the Mercator projection, the compromises by Miller and Robinson, and the "new" (equivalent) projection by Peters.

Oglilby, 1675

American Automobile Association, 1970

BRANIFF'S AIR MAP FOR KANSAS CITY.
Non-stop and thru service.

Braniff Airlines, c. 1980

Riasz, 1958

ST. JOHNSBURY & LAMOILLE COUNTY R.R.

1953

The Direct Bridge Route between Maine, New York State and the Midwest

RETAIL SALES
1948

Harris, 1948

Boston Metropolitan
Transportation
Authority, c. 1980
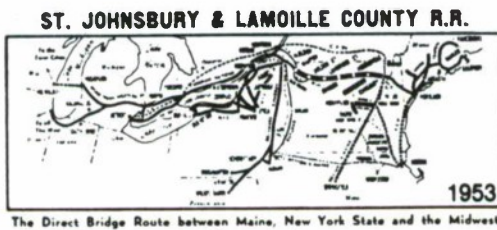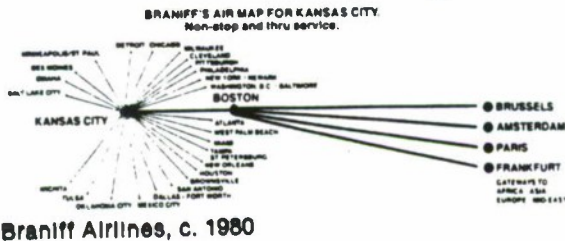
Isidore of Seville, 620 (1472)

Figure 8.– Topological transformations. Two road maps, three centuries apart. Maps from an airline, a railroad line, and a rapid transportation system—with varying levels of schematic development. Two examples of cartograms—with areas on the maps proportional to statistical values (Population by Riasz, and retail sales by Harris). The oldest printed map, a schematic view of the world drawn originally by a seventh-century Christian scholar—a graphic display of the world derived from the Bible.
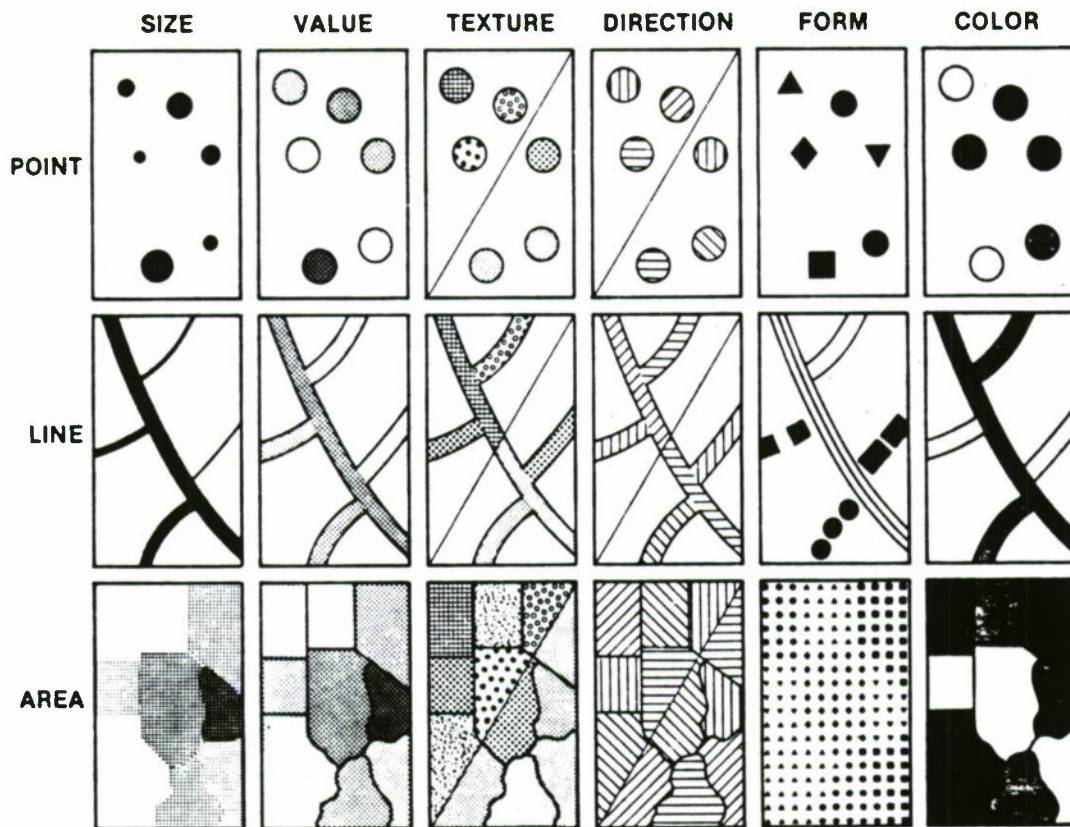
Figure 9.– The visual variables (after the work of Bertin).
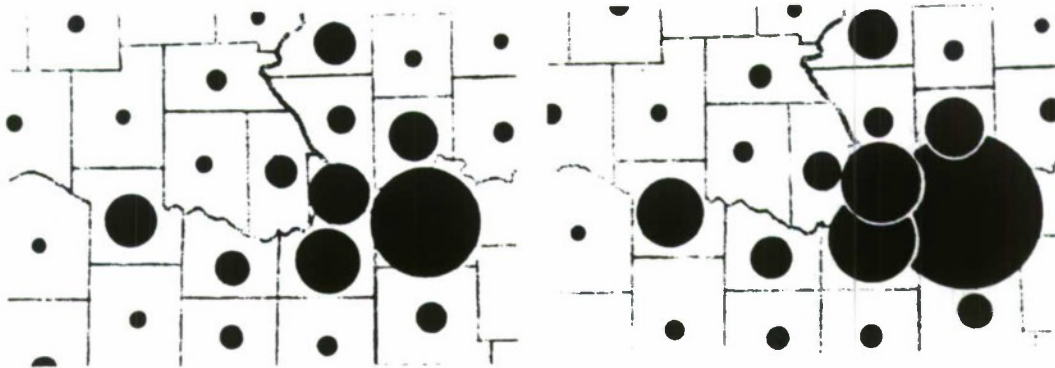
Figure 10.– Two maps, prepared to represent county populations in Kansas and Missouri. The circles on the map at the left are scaled so that their physical areas are directly proportional to the county populations. In the map at the right, the circles have been rescaled so that their size differences are increased, an effort to overcome the "natural" tendency of most map readers to underestimate size differences of point symbols.
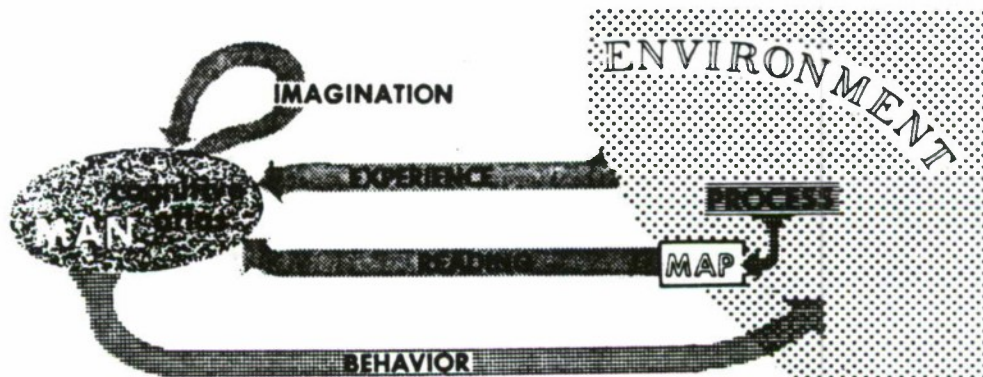


Figure 11.– Model of the cartographic process.

# MANIPULATIVE CONTROL

# MULTI-AXIS CONTROL OF TELEMANIPULATORS

G.M. McKinnon[1] and Ron Kruk[2]
CAE Electronics Ltd.
Montreal, Canada

## ABSTRACT

This paper describes the development of multi-axis hand controllers for use in telemanipulator systems. Experience in the control of the SRMS arm is reviewed together with subsequent tests involving a number of simulators and configurations, including use as a side-arm flight control for helicopters. The factors affecting operator acceptability are reviewed.

## INTRODUCTION

The success of in-orbit operations depends on the use of autonomous and semiautonomous devices to perform construction, maintenance and operational tasks. While there are merits to both fully autonomous and man-in-the-loop (or teleoperated) systems, as well as for pure extravehicular activity (EVA), it is clear that for many tasks, at least in early stages of development, teleoperated systems will be required.

This paper reviews some experience gained in the design of the human-machine interface for teleoperated systems in space. A number of alternative approaches have been proposed and evaluated over the course of the work described, and some basic design principles have evolved which may appear mundane or obvious after the fact, but which nevertheless are critical and often ignored.

One key design objective in the implementation of human-machine interfaces for space is that of standardization. Astronauts should naturally and comfortably interpret their input motions in terms of motions of the manipulator or task. This "transparency" is achieved by careful design to ensure that task coordinates and views are always presented in a clear, unambiguous and logical way, and by ensuring that standardized input devices are used in standardized modes. If conventions are established and systematic modes of control are respected, training time is reduced and effectiveness and performance are improved. The end objective in the design of displays and controls for telemanipulators is to establish a "remote presence" for the operator.

## THE SRMS SYSTEM

A number of manual control input devices have been used in space over the years. For the most part these devices were designed as flight controls for the various satellites and modules

---

which have flown. The first truly robotic control device was that used on the SRMS or CANADARM system of the Space Shuttle. The control interface in this case consisted of two three-degree-of-freedom devices used in conjunction with a displays and controls panel, CCTV visual feedback from cargo bay and arm-mounted cameras, augmented by limited direct viewing. A Translational Hand Control (THC) allowed the astronaut to control the end point of the arm in the three rectilinear degrees of freedom with the left hand, and a Rotational Hand Control (RHC) was used in the right hand to control rotational degrees of freedom.

The THC was designed specifically for the SRMS application by CAE Electronics, while the RHC was a modified version of the Shuttle flight control produced by Honeywell. The geometry and overall configuration of the RHC was thus predetermined and was not matched to the task. The device does not have the single centre of rotation which is considered by the authors to be an advantage in generalized manipulator control. The RHC differed from the flight control version in several ways:

- The forces and travels were modified to reflect task requirements.

- Auxiliary switches and functions were changed to comply to task requirements. In fact all auxiliary switches were located on the RHC -COARSE/VERNIER, RATE HOLD and CAPTURE RELEASE.

- A switch guard was added to CAPTURE/RELEASE to prevent inadvertent release of a payload.

- Redundant electronics were eliminated in view of the reduced level of criticality.

The THC differed from the RHC in that it incorporated rate-dependent damping through the use of eddy current dampers driven by planatary gears. A hand index ring was added to the THC after initial evaluations of prototype units. The ring provided a reference for position and led to the use of the device as a fingertip control, whereas the RHC with its larger hand grip was clearly a hand control. Force levels and gradients on the THC were low, and the rate dependent damping enhanced the smooth feel of the device. The x and y inputs of the THC were not true translations, but an effort was made to optimize a linkage in the available space to reduce the curvature due to a displaced pivot point.

The SRMS system has proven to be operable but not optimal. With training, astronauts can become proficient in performing required tasks. In general, however, the tasks must be carefully programmed and significant training and practice is required before an astronaut feels comfortable with the system. Even with training, the skill of the astronaut is still a limiting factor on system capability. Tasks requiring coordinated or dextrous motions are difficult to achieve.

While there is no hard data to compare alternatives, the shortcomings of the SRMS design in part can be attributed to the limitations of the RHC and THC described above, but mainly to the unfortunate location of the two hand controls and lack of direct correspondence between the axes of the controls and those of the visual displays.

The SRMS system incorporated no force-reflective feedback aside from indications of motor parameters from each joint. Positional feedback of the end point is strictly visual–either direct viewing or through CCTV. The axes of the presented display depend on the view selected: direct,

cargo bay or arm-mounted camera. Control is in the resolved rate mode. In the case of a large-scale arm such as the SRMS, a master slave or indexed position mode is not suitable because of scaling problems.

Figure 1 shows a simulation of the SRMS Displays and Controls System in SIMFAC. The RHC is located to the lower right of the D&C panel and a breadboard model of the THC to the upper left. The CCTV displays are to the right and the direct viewing ports are overhead and immediately above the D&C panel.

## MULTI-AXIS STUDY

Following the design of the SRMS system, the authors conducted a study of multi-axis controls (1). The purpose of the study was to determine the feasibility of controlling six degrees of freedom with a single hand control. According to the guidelines laid down for the study, mode changes were to be avoided so that coordinated control was required simultaneously in all axes. No specific application was defined; however, the controller was to be usable either to fly a spacecraft or to "fly" the end point of a manipulator.

The study included a review of the literature, observation of available multi-axis controllers, and discussions with experts. Although a prototype device was not required by the contract, one was assembled. Interestingly, the consensus of opinion at the time amongst the knowledgeable community was that coordinated control in six axes was desirable, but probably not feasible.

A number of six-degree-of-freedom controls were reviewed. The most notable were devices with force feedback operated in the indexed position mode. A prototype laboratory version was developed by R. Skidmore at Martin Marietta and evaluated in various dynamic and graphic simulations. A similar design and evaluation was done at Jet Propulsion Laboratories by A. Bejczy (2). These devices were both unsuitable in design for implementation in a mature control system, but permitted laboratory evaluation of force characteristics, displacements, and interactions with visual feedback. Another approach was developed by D. Whitney at the Draper Laboratory. This was elegantly designed from the mechanical viewpoint, but difficult to use due to the absence of tactile feedback.

This study uncovered no mature or workable concept for a six-degree-of-freedom controller and a lot of skepticism amongst practitioners as to the feasibility of implementing more than four degrees of freedom. A more recent study of hand controls was done by Brooks and Bejczy (3).

## DEVELOPMENT PROCESS

At the conclusion of the study, in spite of the climate of skepticism, the authors felt that there was no reason why a well-coordinated, six-degree-of-freedom controller could not be designed. Experiments with a variable-geometry test rig demonstrated that the only way to avoid inherent cross-coupling between axes, achieve the ability to make discrete inputs where required, and still have a direct correlation between control inputs and resulting action was to center all axes at a single point positioned at the geometric center of the cupped hand. In this way, control of the end

point related to hand motions. Alignment of controller axes in a logical way to the axes of visual displays was also considered essential.

One initial concern was the issue of isometric (purely force) versus displacement control. An isometric controller is rugged and easily constructed from a mechanical standpoint. Unfortunately, the concept leads to overcontrol, particularly in stressful situations, because of the lack of proprioceptive indication of input commands. In some situations operators tend to saturate the controller to the extent that they quickly suffer fatigue. While there may be tasks in which isometric control is adequate and acceptable, in general the addition of displacement with suitable breakout gradients and hard-stop positions improves performance. For this reason, most manual controls designed on the isometric principle have been modified to include compliance.

Initial designs by the authors were based on the use of force transducers to generate input signals. The controls were designed to allow for the inclusion of compliance and adjustable force characteristics, although the device could also be configured for isometric operation in all axes. It was quickly established that some compliance was advantageous. Since there was always significant displacement, the force transducers were replaced by position transducers, thus permitting the use of rugged, compact, noncontact, optical position sensors and eliminating the tendency to generate noise signals due to vibration or shock. In addition, a purely position system made it easier to eliminate cross-coupling between axes when pure motions in a single axis were required.

An intermediate step of isometric translational axes and displacement in rotation, a so-called "point and push" approach, was unsuccessful because of the problems described above in the isometric axes.

In the final analysis, a prototype design was constructed which included significant displacement in all six axes. The prototype unit is shown in figure 2.

## PROTOTYPE DESIGN

The design concept was to ensure that all six axes pass through a single point. The mechanical components and transducers for the rotational axes were mounted within a ball. The ball in turn was mounted on a stick which was free to translate in three mutually orthogonal axes. All axes had appropriate breakout forces, gradients and stop-force characteristics generated by passive components. The output of the device was a position signal sensed by optical transducers. No additional rate-dependent damping was included. While rate-dependent damping does enhance the "feel" of the controller, the additional mechanical complexity is probably not justified.

The relationship between breakout forces and gradients is task-dependent. In general, the breakouts should be sufficient that pure inputs can be generated easily in a single axis; however, breakouts do have a negative impact on controllability for small coordinated movements in multiple axes simultaneously.

Various handgrip shapes were investigated, but with the emergence of the coincident axis concept as previously described, there was a fundamental need to provide a face perpendicular to the direction of commanded motion. The other prime requirement was a shape which ensured the correct positioning of the hand relative to the geometric center of the system. The natural solution

was a sphere. As development of the mechanism and sensing systems progressed, the ball size was reduced to its present configuration. This approximates to the size of a baseball, and has shown to be comfortable for bare-handed, gloved, and pressure-suited operation.

Several derivatives of the basic design evolved for special applications. A bang-bang device was configured for tests on the MMU simulator. A four-axis (three rotations on a vertical purely rate-dependent damped linear axis) model was evaluated for flight control in helicopters. In some configurations a protuberance was added to provide a tactile cue for orientation. Auxiliary switches were added on this protuberance.

## TEST AND EVALUATION

To date a number of tests have been carried out. It is difficult to compare data between tests since different tasks and performance metrics were used. In general, though, subjective ratings and measures of performance were consistent and some basic design principles were established. Tests performed were

### Johnson Space Flight Center

Initial tests were performed using the controller to control computer graphic representations of docking tasks.

Subsequent tests were also made using the full-scale mockup of the SRMS arm (MDF). Comparisons were made between the conventional SRMS (two three-degree-of-freedom controllers) configuration and the single six-axis device. NASA human factors personnel, technicians and astronauts participated in the tests.

### Martin Marietta

The controller was evaluated with computer graphics representations of docking maneuvers.

Astronaut evaluations of a bang-bang configuration were done on the MMU simulator. Tests were performed for operation in pressurized space suits, as shown in figure 3.

### Marshall Space Flight Center

A six-axis controller was used to control a six-axis arm as shown in figure 4. The system has been operated over the past 2 years with a variety of operators and tests.

## Grumman

Tests were carried out using two six-degree-of-freedom controllers to control two six-degree-of-freedom dextrous manipulators as shown in figure 5. Comparisons were done with master/slave control in the same environment.

Tests were carried out using the six-degree-of-freedom controller with the LASS simulator for various "cherry picker" tasks.

## National Aeronautical Establishment

Four-axis versions of the design were installed and flown in a variable-stability helicopter as shown in figure 6. Evaluations were performed by numerous military and civilian pilots, including test pilots from major airframe manufacturers. Cooper-Harper ratings were recorded for a variety of maneuvers at various levels of control augmentation. Results were comparable to conventional controls. For the most part flight tests were performed by highly experienced pilots.

It should be noted that, in the case of the four-axis version, the use of a relatively conventional handgrip superimposed on the ball was possible while respecting the principle of a single centre. The addition of another translation axis with a similar handgrip would introduce cross coupling.

## European Space Agency

A model of the controller has been ordered by ESA for evaluation use in the European Space Program.

## DISCUSSION

Tests to date have demonstrated that six-axis control using a single hand is not only feasible but, providing certain design guidelines are respected, preferable to approaches in which axes are distributed amongst separate controllers. Statements to the effect that six degrees of freedom is too much for one hand ignore the fact that the humans have the ability to make complex multi-axis movements with one hand using only "end point" conscious control. The coordinate transformations required are mastered at an early age and the inverse kinematics are resolved with no conscious effort. To operate a system using two separate three-axis controllers requires a conscious effort on the part of the operator, thereby increasing his or her work load. The operator requires considerable training and practice with a 2 x 3 axis system before achieving the same level of control as is immediately possible with the single six-axis device. NASA experience has shown that the weeks of training necessary for the former become less than 30 sec for the latter. While the guidelines have been verified only in specific environments for specific tasks, the authors feel confident in making the following statements:

1. A proportional displacement controller will provide improved performance and in many cases more relaxed control than an isometric device. Performance with isometric devices varies more between individual subjects than that with displacement control.

2. Force gradients and characteristics should be correlated to the task being performed. There may be a justification for standardizing force characteristics and controller configurations for all space-related equipment to ensure commonality and to reduce training requirements.

3. An obvious and consistent orientation between controller axes and those of visual feedback displays is essential. This is an area where standardization between tasks and systems is a key element. A single controller design would be suitable for all applications, provided that basic axis orientation and control mode standards are maintained.

4. The use of force-reflecting feedback has not been evaluated by the authors, although a program is under way to investigate some unique and novel approaches. In general, direct force feedback is useful only in a system with high mechanical fidelity. In the presence of abrupt non-linearities such as stiction or backlash and particularly transport lag force, feedback can in fact be detrimental in excess of 100 msec.

5. For some tasks with some manipulators, a master/slave system can provide equal or superior performance to that of a manual control in resolved-rate mode. Resolved rate is, however, universally applicable and can provide a standardized approach for virtually all manipulator or flight-control tasks.

6. In tasks in which lag exceeds 1 sec, it may be assumed that real-time interactive control in the strict sense is not feasible. Providing physical relationships are stable or static, a reconstructive mode using generated graphics for a "prehearsal" of manipulator movement may be used, stored in memory, then activated. When lags are 100 msec or less, resolved-rate control may be used to directly control the end-effector (position control is inadequate when any substantial excursion may be required). The lag regime between 100 msec and 1 sec causes difficulty because there is a tendency to compensate for delay or system instability (e.g., arm-flexing modes) with more complex drive and "prediction" algorithms. Our experience thus far is that the simplest control algorithm which permits stable response generally provides the best performance.

In conclusion, tests have shown that six-degree-of-freedom controllers can be used naturally and effectively to control tasks requiring dexterity and coordination.

# REFERENCES

1. Lippay, A.L., McKinnon, G.M., King, M.L., Multi-axis Hand Controllers, a State of the Art Report, Proceedings of Seventeenth Annual Conference on Manual Control, Los Angeles, June 1981.

2. Bejczy, A. K., Handlykon, M., Experimental Results with a Six Degree of Freedom Force Reflecting Hand Controller, Proceedings of Seventeenth Annual Conference on Manual Control; Los Angeles, June 1981.

3. Brooks, T. L., Bejczy, A. K., Hand Controls for Teleoperation, VPL Publication 85-11, California Institute of Technology, Pasedena, March 1985.
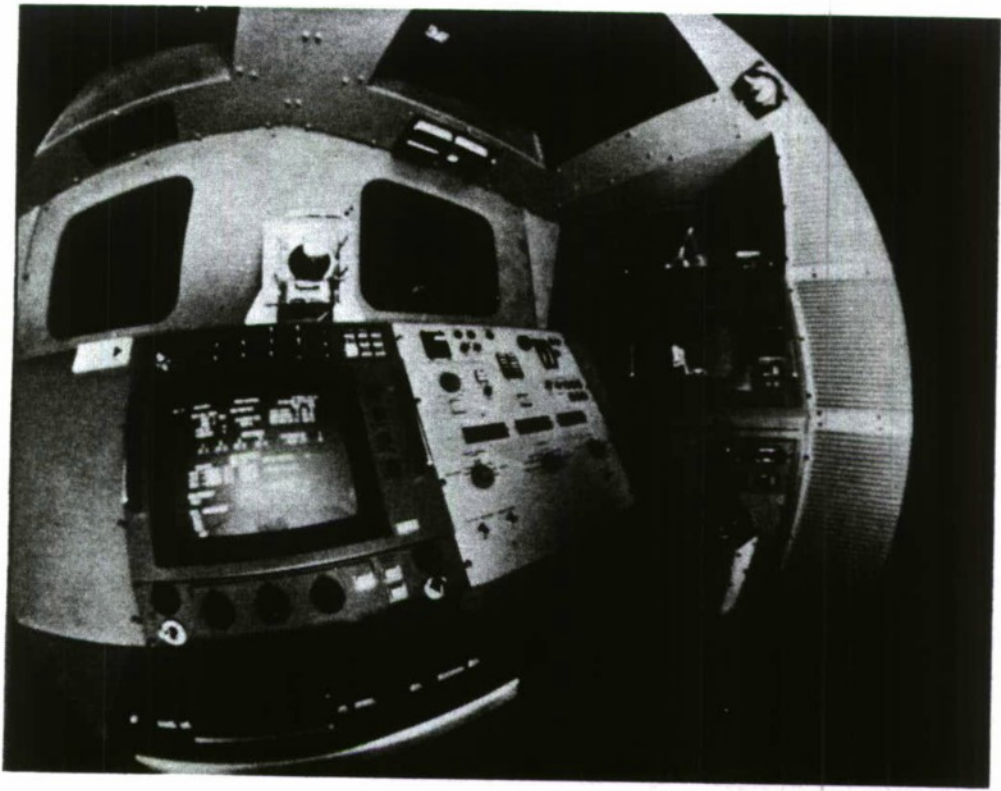
Figure 1.- SIMFAC.
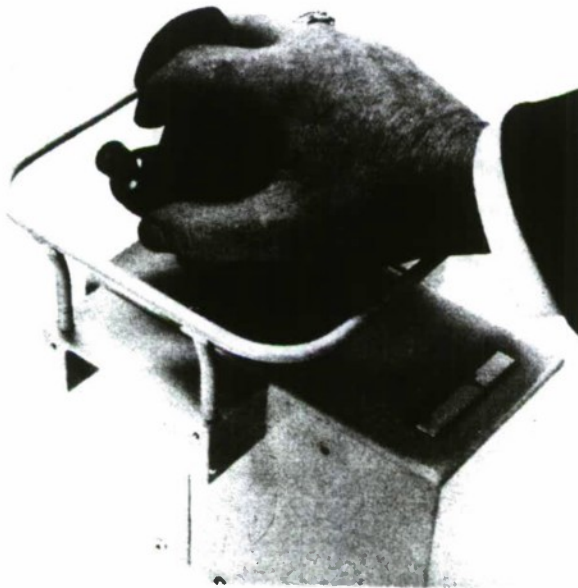


Figure 2.- Six-degree-of-freedom prototype.
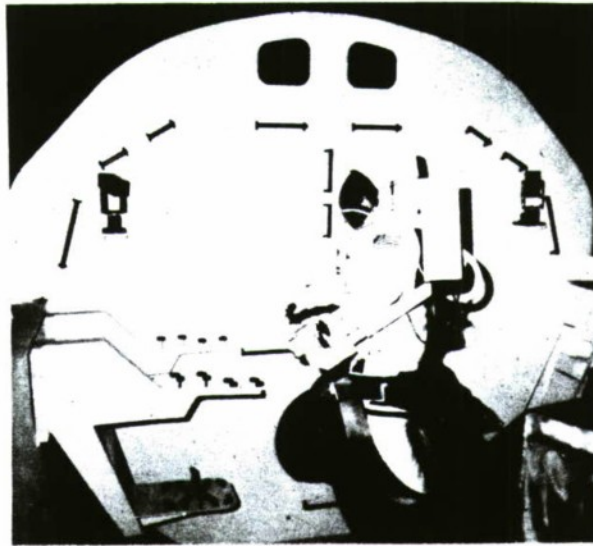
Figure 3.- MMU tests at Martin Marietta.



Figure 4.- Control of robot at Marshall Space Flight Center.

Figure 5.- Simultaneous control of two arms.



Figure 6.- Four-degree-of-freedom controller installed in helicopter.

# TELEPRESENCE, TIME DELAY, AND ADAPTATION

Richard Held
Massachusetts Institute of Technology, Cambridge, Massachusetts

Nathaniel Durlach
Massachusetts Institute of Technology, Cambridge, Massachusetts
and
Boston University, Boston, Massachusetts

## INTRODUCTION

Displays, which are the subject of this conference, are now being used extensively throughout our society. More and more of our time is spent watching television, movies, computer screens, etc. Furthermore, in an increasing number of cases, the observer interacts with the display and plays the role of operator as well as observer. To a large extent, our normal behavior in our normal environment can also be thought of in these same terms. Taking liberties with Shakespeare, we might say that "all the world's a display and all the individuals in it are operators in and on the display."

Within this general context of interactive display systems, we begin our discussion with a conceptual overview of a particular class of such systems, namely, teleoperator systems. We then consider the notion of telepresence and the factors that limit telepresence, including decorrelation between the (1) motor output of the teleoperator as sensed directly via the kinesthetic/tactual system, and (2) the motor output of the teleoperator as sensed indirectly via feedback from the slave robot, i.e., via a visual display of the motor actions of the slave robot. Finally, we focus on the deleterious effect of time delay (a particular source of decorrelation) on sensory-motor adaptation (an important phenomenon related to telepresence).

## I. TELEOPERATOR SYSTEMS

A schematic outline of a highly simplified teleoperator system is presented in figure 1. As pictured, the major components of a teleoperator system are a human operator, a teleoperator station (or "suit"), a slave robot, and an environment which is sensed and acted upon by the slave robot. As indicated by the arrows flowing from left to right, sensors on the slave robot are stimulated by interaction with the environment, the outputs of these sensors are displayed in the teleoperator station to the sensors of the human operator, and the received information is then transmitted to higher centers (brain) within the human operator for central processing. As indicated by the arrows flowing from right to left, the central processing results in motor responses by the human operator which are detected in the teleoperator station and used to control motor actions by the slave robot. The upward flowing arrows depict the role played by the motor system (at both the slave robot and human operator levels) in controlling the sensors and therefore the flow of information from environment to brain.

The normal situation in which the human interacts directly with the environment can be pictured as a special case of the teleoperator situation by ignoring the teleoperator station and identifying the slave robot's sensors and effectors with those of the human operator. Similarly, imaginary or virtual environments can be pictured in terms of the teleoperator situation by retaining the human operator and teleoperator station, but replacing the real environment and slave robot by a computer simulation. Finally, robotic systems can be realized by replacing the human operator and teleoperator station by an automatic central processor, and interpolations between teleoperator systems and robotic systems can be realized by assigning lower-level control functions to automatic processing and higher-level control functions (supervisory control) to the human operator.

Note also that the sensor and effector channels need not be restricted in the manner illustrated in Fig. 1. Not only are there many cases in which the visual channel pictured would be paralleled by an auditory channel, but for certain purposes the slave robot might also include sensors for which the human has no counterpart (e.g., to sense infrared energy or magnetic fields). Furthermore, on the response side, the teleoperator station might detect and exploit responses other than simple motor actions. For example, it might be useful for certain purposes to measure changes in skin conductivity, pupil size, or blood pressure.

In general, the purpose of a teleoperator system is to augment the sensory-motor system of the human operator. The structure of the teleoperator system will depend on the specific augmentation envisioned, as well as on the technological limitations. A continuum that relates directly to the issue of telepresence considered below concerns the extent to which the structure of the slave robot is the same as that of the teleoperator. At one extreme are systems meant simply to transport the operator to a different place. In the ideal version of such a system, the slave robot would be isomorphic to the operator and the various sensor and effector channels would be designed to realize this isomorphism. In a closely related set of systems, the basic anthropomorphism is preserved, but the slave robot is scaled to achieve, for example, a reduction of size or magnification of strength. At the opposite extreme are systems involving radical structural transformations and highly non-anthropomorphic slave robots. In these systems, there is no simple correspondence between slave robot and human operator, and the design and organization of the sensor and effector channels generally becomes very complex and difficult to optimize, even at the abstract conceptual level. General reviews of teleoperation and teleoperator systems can be found in Johnsen and Corliss, 1974, and Vertut and Coiffet, 1986.

## II. TELEPRESENCE

Although the term "telepresence" is often used in discussions of teleoperation, it never has been adequately defined. According to Akin et al. (1983), telepresence occurs when the following conditions are satisfied:

> "At the worksite, the manipulators have the dexterity to allow the operator to perform normal human functions. At the control station, the operator receives sufficient quantity and quality of sensory feedback to provide a feeling of actual presence at the worksite."

A major limitation of this definition is that it is not sufficiently operational or quantitative. It does not specify how to measure the degree of telepresence. Also, as indicated by the phrase

"perform normal human functions" in the first sentence, it fails to address the issue of telepresence for systems that are designed to transform as well as transport and to perform abnormal human functions.

Independent of the precise definition of telepresence, why should one care about telepresence? What is it good for? Certainly, there is no theorem which states that an increase in telepresence necessarily leads to improved performance. In our opinion, a high degree of telepresence is desirable in a teleoperator system primarily in situations when the tasks are wide-ranging, complex, and uncertain, i.e., when the system must function as a general-purpose system. In such situations, a high degree of telepresence is desirable because the best general-purpose system known to us (as engineers) is us (as operators). In a passage that is relevant both to this issue and to the definition of telepresence, Pepper and Hightower (1984) state the following:

> "We feel that anthropomorphically-designed teleoperators offer the best means of transmitting man's remarkably adaptive problem solving and manipulative skills into the ocean's depths and other inhospitable environments. The anthropomorphic approach calls for development of teleoperator subsystems which sense highly detailed patterns of visual, auditory, and tactile information in the remote environment and display the non-harmful, task-relevant components of this information to an operator in a way that very closely replicates the pattern of stimulation available to an on-site observer. Such a system would permit the operator to extend his sensory-motor functions and problem solving skills to remote or hazardous sites as if he were actually there."

In addition to the value of telepresence in a general-purpose teleoperator system, it is likely to be useful in a variety of other applications. More specifically, it should enhance performance in applications (referred to briefly in Sec. I) where the operator interacts with synthetic worlds created by computer simulation. The most obvious cases in this category are those associated with training people to perform certain motor functions (e.g., flying an airplane) or with entertaining people (i.e., providing imaginary worlds for fun). Less obvious, but equally important, are cases in which the system is used as a research tool to study human sensorimotor performance and cases in which it is used as an interactive display for data presentation (e.g., Fisher, 1987; Bolt, 1984).

An important obstacle at present to scientific use of the telepresence concept is the lack of a well-defined means for measuring telepresence. It should not only be possible to develop subjective scales of telepresence (using standardized scale-construction techniques), but also to develop tests, both psychological and physiological, to measure telepresence objectively. For example, some test based on the "startle response" might prove useful. Certainly, such a test could distinguish reliably between different degrees of realism in the area of cinematic projection. Also, of course, given both some subjective scales and some objective tests, it would be important to study the relations among the two types of measures.

Beyond questions related to the definition and measurement of telepresence, the core issue is how one achieves telepresence. In other words, what are the factors that contribute to a sense of telepresence? In fact, what are the essential elements of just plain "presence?" Or alternately, looking at the other side of the coin, how can the ordinary sense of presence be destroyed (short of damaging the brain)?

Given the vague and qualitative character of definitions and estimates of telepresence, it is not surprising that there is no scientific body of data and/or theory delineating the factors that underlie telepresence. Our remarks on this topic thus make substantial use of intuition and speculation, as well as extrapolation from results in other areas.

Sensory factors that contribute to telepresence include high resolution and large field of view. Obviously, reduction of input information either by degraded resolution or restricted field of view will interfere with the extent to which the display system is transparent to the operator. Perhaps these two variables are tradeable in the sense that the effective parameter in determining the degree of telepresence is the number of resolvable elements in the field, or, equivalently, for fields with uniform resolution over the field, Area of Field/Area of Resolvable Element. Also important, of course, is the consistency of information across modalities: the information received through all channels should describe the same objective world (i.e., should be consistent with what has been learned through these channels about the normal world during the normal development process). In addition, the devices used for displaying the information to the operator's senses in the teleoperator station should, to the extent possible, be free from the production of artifactual stimuli that signal the existence of the display. Thus, for example, the visual display should be sufficiently large and close enough to the eyes to prevent the operator from seeing the edges of the display (or anything else in the teleoperator station, including the operator's own hands and body). At the same time, the display should not be head-mounted in such a way that the operator is aware of the mounting via the sense of touch. Clearly, attempting to satisfy both of these constraints simultaneously is a very challenging task.

Motor factors necessary for high telepresence involve similar issues. Perhaps the most crucial requirement is to provide for a wide range of sensorimotor interactions. One important category of such interactions concerns movements of the sensory organs. It must be possible for the operator to sweep the direction of gaze by rotating the head and/or eyeballs and to have the visual input to the retinas change appropriately. This requires using a robot with a rotating head, the position of which is slaved to the position of the operator's head. The desired result can then be achieved in two ways, depending upon whether the system is designed to have the position of the robot's eyeballs (1) fixed relative to the the robot's head (e.g., pointing straight ahead) or (2) slaved to the position of the operator's eyeballs in the operator's head. In the first case, appropriate results can be obtained using binocular images that remain fixed relative to the operator's head position during eyeball scanning. In the second case, the positions of the projected images must be slaved to the position of the operator's eyeballs. If they were instead held fixed, then whenever the operator's eyeballs were rotated, the projected images would rotate. For example, if the operator's eyeballs were rotated to look at an object whose images were on the right side of the projection screens, the slave robot's eyeballs would rotate to the right, the images of the object in question would move to the center of the two screens, and these images would then be sensed to the left of the foveal region. In order to eliminate this problem, the projected images would also have to be rotated to the right. In other words, if the position of the robot eyeballs are slaved, the position of the projected images must also be slaved. To the best of our knowledge, no such system has yet been developed (although monitoring of operator eyeball position is being used to capitalize on reduced resolution requirements in the peripheral field in the pursuit of reduced bandwidth).

Another category of sensorimotor interactions that is essential for high telepresence concerns movements of viewed effectors. It must be possible for the operator to simultaneously move his/her hands (receiving the internal kinesthetic sensations associated with these movements) and

see the slave robot hands move accordingly. Also, as with the sensory display, the devices used in the teleoperator station to detect and monitor the operators movements should, to the extent possible, be undetectable to the operator. The more the operator is aware of these devices, the harder it will be to achieve a high degree of telepresence. An amusing picture that is addressed to the issue of viewing one's own effectors, or more generally, one's own body parts, and that is of some historical interest, is shown in figure 2 (Mach, 1914).

The most crucial factor in creating high telepresence is, perhaps, high correlation between (1) the movements of the operator sensed directly via the internal proprioceptive/kinesthetic senses of the operator and (2) the actions of the slave robot sensed via the sensors on the slave robot and the displays in the teleoperator station. Clearly, the destruction of such correlation in the normal human situation (in which the slave robot is identified with the operator's own body) would destroy the sense of presence.

In general, correlation will be reduced by time delays, internally generated noises, or non-invertible distortions that occur between the actions of the operator and the sensed actions of the slave robot. How these variables interact, combine, and trade in limiting telepresence and teleoperator performance is a crucial topic for research. In sec. III, we look more closely at the effects of one of these variables, namely, time delay.

Note also that telepresence will generally tend to increase with an increase in the extent to which the operator can identify his or her own body with the slave robot. Many of the factors mentioned above (in particular, the correlation between movements of the body and movements of the robot) obviously play a major role in such identification. Additional factors, however, may also be important. For example, it seems plausible that identification, and therefore telepresence, would be increased by a similarity in the visual appearance of the operator and the slave robot.

Finally, it is important to consider the extent to which telepresence can increase with operator familiarization. Even if the system is designed merely to transport rather than to transform, it will necessarily involve a variety of transformations that initially limit the sense of telepresence. A fundamental topic for research concerns the extent to which such limitations can be overcome by appropriate exposure to the system and development of appropriate models of the transformed world, task, self, etc. (through adaptation, training, learning, etc.). Figure 3 illustrates schematically how the internal dynamics of the operator are originally established and may be altered over time when interaction with the world is transformed. The representation (in brain) of the operator's interaction with the world is an important factor in the sense of presence. The operator identifies his or her own actions as such in accord with the concomitant sensory changes. Loss of such concomitance may reduce the sense of presence. But an updating of the internal model may promote the recovery of a lost sense of presence within that world. The figure shows how the motor command originating in the central nervous system (CNS) activates the musculature which in turn causes sensory changes which feed back to the CNS. The comparator is designed to receive a feed-forward signal from the internal model, which derives from past experience and anticipates the consequences of activity based upon that previous experience. That signal is then compared with the contemporary consequences of action. Any transform in the feedback loop will alter the expected feedback and be discrepant with the feedforward signal. In that event, the discrepant signal may be used to update the world model and lead to more accurate anticipations of action and an improved sense of presence.

## III. TIME DELAYS AND ADAPTATION

Time delays between action of the teleoperator and the consequences of these actions as realized on the displays in the teleoperator station can arise from a variety of sources, including the transmission time for communication between the teleoperator station and the worksite and the processing time required for elaborate signal-processing tasks. Independent of the causes, it is clear that such feedback delays degrade both telepresence and performance. Research on the effects of time delays on manual tracking and remote manipulation and on methods for mitigating these effects are discussed in a variety of sources (e.g., Adams, 1962; Arnold and Braisted, 1963; Black, 1970; Ferrell, 1965, 1966; Johnsen and Corliss, 1971; Kalmus, Fry, and Denes, 1960; Leslie, 1966; Leslie, Bennigson, and Kahn, 1966; Levison, Lancraft, and Junker, 1979; Pennington, 1983; Pew, Duffenbach, and Fensch, 1967; Poulton, 1974; Sheridan, 1984; Sheridan and Ferrell, 1963,1967,1974; Sheridan and Verplank, 1978; Starr, 1980; Wallach, 1961; Wickens, 1986). Of particular interest has been the development of systems that combat the effects of time delay through judicious supplementation of human teleoperation by automatic processing (involving predictive models and use of the human operator for supervisory control).

The particular effect of time delay on which we shall focus in the remainder of this paper is the effect on sensory-motor adaptation. As suggested at the end of the last section, the degree of telepresence that can be achieved with a given system depends ultimately on the extent to which the operator can adapt to the system.

Basic demonstration of adaptation was discussed by Helmholtz in his Physiological Optics (Helmholtz, 1962). In the typical experiment, the subject wears prism spectacles over his or her eyes which optically shift the apparent location of objects seen through them. When the subject reaches for a seen target without correction (open loop), the termination of his or her reach will obviously be in error by an amount approximating the apparent displacement of the target produced by the prism. Correction of a reach can be prevented in one of two ways. If the subject ($\underline{S}$) is required to make a rapid ballistic movement of his or her hand to the target, the duration of hand travel is too short to allow correction. However, if both target and hand are visible at the termination of the reach, the error may be noted by $\underline{S}$ and subsequent reaches corrected. Alternatively, the target may be presented in a location where the hand may reach but not be seen. Following the initial measurements of reaching accuracy, the subject views either his or her hand or a surrogate for it through the prisms for a period of time called the exposure period. During that period he may or may not receive visual information concerning the error of the reaching. Following the exposure period a second measure is obtained of the accuracy of open loop reaching for visible targets. The result is generally a decrease of error from that of the initial localizations in a direction which indicates correction for the presence of the prism displacement. Further open-loop measurements may be made with the prisms removed, in which case the error of reaching for a target increases. This increased error shows that the shift in localization is not dependent upon the presence of the prisms, but is a more generalized change in eye-hand coordination adaptive for the presence of the prisms.

Some sort of adaptive process occurs during the exposure period which compensates for the error introduced by the prism. Information available during the exposure period produces an update of the internal model of the visuospatial coordinates which are anticipated as the goal of reaching for the target. The nature of the necessary and sufficient information required for adaptation, and of the subsystems that actually adapt, has been the subject of much debate and

experimentation (Welsh, 1978, 1986). It appears that any of a number of sources of information about the transformed relation between the seen position of the hand and its location as known through other information may serve to produce adapation. One such source of information is the error seen when reaching for targets. When the reaching subject can see the error, he or she is bound to correct for it by a process of which he or she is usually quite conscious. Among other cognitive factors, knowledge of the optical effects of the prism may enhance adaptive responses. Active movement of the arm which produces visual feedback enhances adaptation, perhaps by sharpening the sense of position of bodily parts. More interesting from several points of view is the adaptive process which occurs during exposure when visible error feedback appears to be absent. For example, subjects adapt while looking through the prism at only a luminous spot fixed to the hand in an otherwise dark field. The spot moves with the hand, but when no other targets or even visible landmarks are present, there can be no explicit visible error. There may, however, be a discrepancy with the expectations based upon the concomitance of visual location of the hand with its non-visually sensed position. But this condition raises a further question. If the subject sees only a luminous spot on the hand as it moves, how does the nervous system identify this spot with the sensed positions of the hand? Aside from cognitive factors, we must hypothesize that the movements of the visible spot concomitant with the sensed movements of the hand allow this identification. The problem then becomes one of correlation between signals. Moreover, we recognize that this form of identification may well be a basis for establishing presence itself. This realization led to the following experiment.

The experiment concerns the effect of time delay on adaptation of eye-hand coordination to prism displacement. Changes in the seen position of the hand are delayed during a period of exposure between test and retest. For a given exposure, the delay is fixed, but over a series of exposures, the delay is varied. The question we asked was: What are the effects of delaying feedback by various amounts on the adaptive process that takes place during exposure with continuous monitoring by the subject of his or her hand movements in a frontal plane? In other words, how much is the effective correlation of identifying signals degraded by delay of visual feedback of varying intervals? In an earlier experiment (Held, Efstathiou, and Greene, 1966), we found that delays as small as 300 msec eliminated adaptation to prism displacement. Consequently, the following experiment incorporated delays of smaller magnitude.

As shown in figure 4, the subject (S) stood at the apparatus. He positioned his head in a holder mounted on top of a light-proof box and looked down through an aperture into a mirror. The mirror reflected the image of a luminous spot, formed on a ground glass screen, which appeared on an otherwise dimly illuminated background. The image originated on an oscilloscope face and was focused on the screen. S's right hand grasped a handle consisting of a short vertical rod located at arm's length beneath the box. The rod was attached to a lightweight roller-bearing arrangement which minimized inertia and friction but restricted hand movements to a region in the horizontal plane. When the hand moved the cursor, sliding contacts were driven along two linear potentiometers aligned at right angles to each other. This movement varied DC signals corresponding to the coordinates of the cursor on the horizontal surface. These signals were applied to the vertical and horizontal channels of the oscilloscope, thereby producing a single spot on the screen, the position and motion of which corresponded to that of the cursor. The optical system (lens and mirror) caused the spot to appear superimposed on the handle of the cursor when neither positional displacements nor temporal delays were introduced. The apparatus could be set to displace the spot 1.5 in. laterally to either the right or the left side. Temporal delays ranging from 20 to 1,000 msec could be introduced in either the lateral or the vertical dimension, or both.

In addition to driving the trace by movements of the cursor, the loop could be opened and the trace spot set to display, one a a time, five stationary visible targets. The target coordinates were determined by applying paired X and Y voltages to the oscilloscope under the experimenter's control. Ss were instructed to set the handle of the cursor so that the top of the vertical rod felt superimposed on the visible target. Ss pressed a switch when they felt that the cursor was correctly positioned and the position was recorded.

Ss were 12 right-handed male college undergraduates with adequate vision and were naive as to the purpose of the experiments. Each S performed six runs separated by rest periods. Each run consisted of six steps:

1. Practice. S was instructed to track the luminous spot with his eyes as he moved the cursor back and forth across the horizontal surface and to change the left-right direction of his hand movement with the beat of a metronome. This beat varied in a 60-sec cycle from 50 to 90 beats/min. Practice lasted a minute or two during which the subject traced the limits of movement of the cursor. He was instructed to avoid hitting the limiting stops during subsequent exposure and target localization, thus eliminating one potential source of information regarding the position of his hand on the surface.

2. Pre-Exposure Localization. S was instructed to look at and localize the apparent positions of each of the five visual targets presented four times in a pseudo-random sequence. The moveable spot was extinguished prior to target presentations and the subject was instructed to move the cursor randomly about the surface before and between target presentations.

3. First Exposure. S performed for 2 min as he did during the practice period. Both positional displacement and delayed visual feedback were introduced. One of six delay conditions, 0, 120, 150, 210, 330, and 570 msecs, was presented during each run. The six delays were presented to each S in a different order; half of the Ss were exposed to the spot laterally displaced in one direction (right or left) during this exposure and half with the same order of delayed conditions, but with the direction of displacement in the opposite direction.

4. First Post-Exposure Localization. Identical to the pre-exposure localization.

5. Second Exposure. Identical to the initial exposure, but with lateral displacement in the opposite direction.

6. Second Post-Exposure Localization. Same as pre-exposure localization.

The results were analyzed by taking the differences between the first and second post-exposure localizations as the primary measure of compensatory shift. These differences tend to be larger and more reliable than those between pre-exposure and post-exposure localizations (Hardt, Held, and Steinbach, 1971).

Four experiments were performed. They were identical except for variations in the exposure procedure. In the first experiment, S tracked the hand-driven spot with his eyes as described above. In the second, S's eyes fixated a dim cross during exposure, thereby precluding tracking of the spot with the eyes. In the third, each S was trained to relax his arm while grasping the cursor and the experimenter moved the cursor in the manner discussed above (passive condition).

The fourth experiment was identical to the second except that two shorter time delays were used, namely, 30 and 60 msec.

The $\underline{S}$'s mean compensatory shifts at various time delays are shown in figure 5. The overall effect of delay in the first experiment (no fixation) is significant. All of the mean shifts are different from zero and all the shifts under delay are significantly less than the shift at zero delay. The results of the second experiment (fixation) did not differ significantly from those of the first, showing that tracking the hand-driven target with the eyes was not a factor in promoting adaptation. While the passive condition of the third experiment reduced the overall level of adaptation, significant adaptation still occurred, and the overall shape of the curve with delay was similar to that of the active conditions. Finally, the effects of very short delays in the fourth experiment did not differ significantly from zero delay, although delays of 120 msec clearly do reduce adaptation. We conclude that delays must exceed 60 msec if they are to be sufficient to reduce adaptation significantly under the conditions of the experiment. For reasons we do not understand, the curves appear to asymptote at 30 to 40% of compensation under zero delay.

It should also be noted that subjective impressions varied strongly with the delay. At the shorter delays (not too far above threshold), the viewed hand seems to be suffering simply a minor lag, as if it were being dragged through a viscous medium. At delays beyond a couple of hundred msec, however, the image seen becomes more and more dissociated from the real hand (i.e., identification, and therefore presence, breaks down).

In general, it is obvious that some degree of identification is necessary in order for adaptation to occur. Moreover, when adaptation occurs, it is obvious that identification increases. Thus, adaptation and identification (and therefore telepresence) must be very closely related. Note, however, that adaptation will fail to occur when either (1) no identification is possible or (2) identification is complete. Thus, tests of adaptation cannot, by themselves, be used to measure identification; other kinds of tests must also be included. Clearly, a precise characterization of the relations between adaptation, identification, and telepresence (or presence) requires further study.

## ACKNOWLEDGMENTS

# REFERENCES

Adams, J. L. (1962). An investigation of the effects of time lag due to long transmission distances upon remote control, Phase II, NASA TN-1351.

Akin, D. L., Minsky, M. L., Thiel, E. D., and Kurtzman, C. R. (1983). Space applications of automation, robotics, and machine intelligence systems, (ARAMIS)-Phase II, Vol. 3: Executive Summary, M.I.T., Contract NAS 8-34381, NASA Marshall Space Flight Center.

Arnold, J. E., and Braisted, P. W. (1963). Design and evaluation of a predictor for remote control systems with signal transmission delays, NASA TN D-2229.

Black, J. H., Jr. (1970). Factorial study of remote manipulation with transmission time delay, M.S. Thesis, Dept. Mechanical Engineering, M.I.T., Cambridge, MA.

Bolt, R. A. (1984). The Human Interface, M.I.T., Lifetime Learning Publications, Belmont, CA (a division of Wadsworth, London).

Ferrell, W. R. (1966). Delayed force feedback, Human Factors, 8, 449-455.

Ferrell, W. R. (1965). Remote manipulation with transmission delay, M.I.T., NASA TN D-2665.

Fisher, S. (1987). Telepresence in dataspace, Proceedings of symposium and workshop on Spatial Displays and Spatial Instruments, NASA and University of California (Berkeley), Asilomar, CA.

Hardt, M. E., Held, R., and Steinbach, M. J. (1971). Adaptation to displaced vision: a change in the central control of sensorimotor coordination, J. Exp. Psychol., 89, 229-239.

Held, R., Efstathiou, A., and Greene, (1966). Adaptation to displaced and delayed visual feedback from the hand, J. Exp. Psychol., 72, 887-891.

Helmholtz, H. Von (1962). Treatise on physiological optics (Vol. III), trans. from the 3rd German (ed., J. P. C. Southall), Dover, New York.

Johnsen, E. G., and Corliss, W. R. (1971). Human factors: applications in teleoperator design and operation, Wiley Interscience, New York.

Kalmus, H., Fry, D. B., and Denes, P. (1960). Effects of delayed visual control on writing, drawing, and tracing, Language Speech, 3, 96-108.

Leslie, J. M. (1966). Effects of time delay in the visual feedback loop of a man-machine system, NASA CR-560.

Leslie, J. M., Bennigson, L. A., and Kahn, M. E. (1966). Predictor aided tracking in a system with time delay, performance involving flat surface, roll, and pitch conditions, NASA CR-75389.

Levison, W. H., Lancraft, R. E., and Junker, A. M. (1979). Effects of simulator delays on performance and learning in a roll-axis tracking task, AFFDL-TR-79-3134, Proc. 15th Ann. Conf. on Manual Control, Dayton, OH.

Mach, Ernst (1914). The analysis of sensations. Open Court Publishing Company, Chicago.

Pennington, J. E. (1983). A rate-controlled teleoperator task with simulated transport delays, NASA TM-85653.

Pepper, R. L., and Hightower, J. D. (1984). Research issues in teleoperator systems, 28th Annual Human Factors Society Meeting, San Antonio, TX.

Pew, R. W., Duffenbach, J. C., and Fensch, J. C. (1967). Sine wave tracking revisited, Trans., IEEE, HFE-8, 130-134.

Poulton, E. C. (1974). Tracking skill and manual control. Academic Press, New York.

Sheridan, T. B. (1984). Supervisory control of remote manipulators, vehicles, and dynamic processes: experiments in command and display aiding, in Advances in Man-Machine Systems Research, Vol. I, (W. B. Rouse, Ed.), pp. 49-137, J.A.I. Press, Greenwich, CT.

Sheridan, T. B., and Ferrell, W. R. (1963). Remote manipulation control with transmission delay, Trans. IEEE, Vol. HFE-4, 25-28.

Sheridan, T. B., and Ferrell, W. R. (1967). Supervisory control of manipulation, NASA SP-144, 315-323.

Sheridan, T. B., and Ferrell, W. R. (1974). Man-machine systems, M.I.T. Press, Cambridge, MA.

Sheridan, T. B., and Verplank, W. L. (1978). Human and computer control of undersea teleoperators, Department of Mechanical Engineering, M.I.T., Technical Report, Engineering Psychology Program, ONR.

Starr, G. P. (1980). A comparison of control modes for time-delayed remote manipulation, Proc. 16th Ann. Conf. on Manual Control, M.I.T., Cambridge, MA.

Vertut, J., and Coiffet, P. (1986). Robot Technology, Teleoperations and Robotics: Evolution and Development (Vol. 3A) and Applications and Technology (Vol. 3B), Prentice-Hall, Englewood Cliffs, NJ.

Wallach, H. C. (1961). Performance of a pursuit tracking task with different time delay inserted between the control mechanism and display cursor, TM 12-61, U.S. Ordnance Human Engineering Laboratory.

Welsh, R. B. (1986). Adaptation of space perception, Ch. 24 in Handbook of Perception and Human Performance, Vol. I (K. R. Boff, L. Kaufman, J. P. Thomas, Eds.) John Wiley & Sons, New York.

Welsh, R. B. (1978). Perceptual Modification, Adaptions to Altered Sensory Environments, Academic Press, New York.

Wickens, C. D. (1986). The effects of control dynamics on performance, Ch. 39 in Handbook of Perception and Human Performance, Vol. II (K. R. Boff, L. Kaufman, J. P. Thomas, Eds.) John Wiley & Sons, New York.
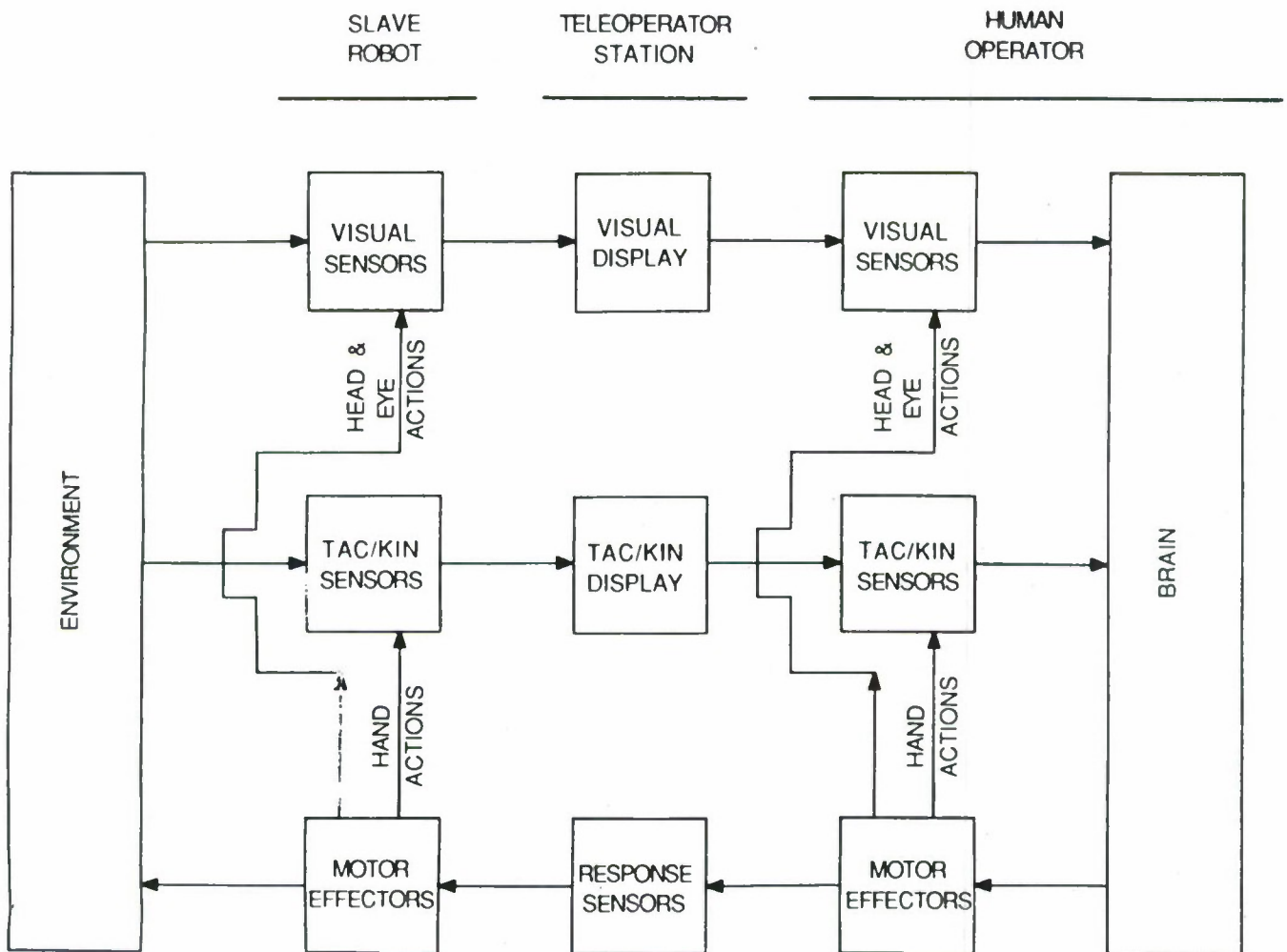
Figure 1.– Schematic outline of teleoperator system.

Figure 2.– Mach observing visible parts of his own body and the surroundings.
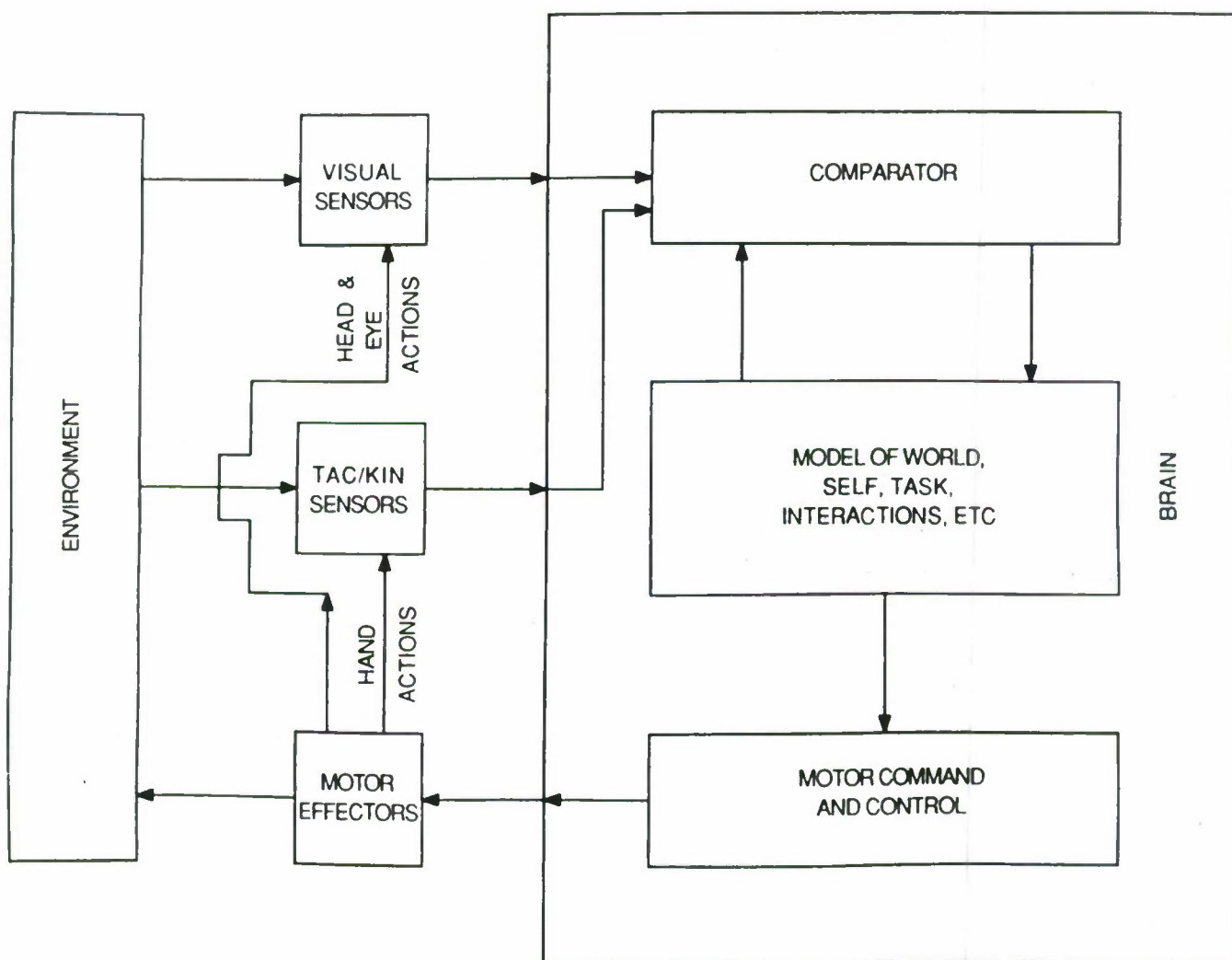
Figure 3.– Information flow and feedback loops involved in actions of the operator in the environment.
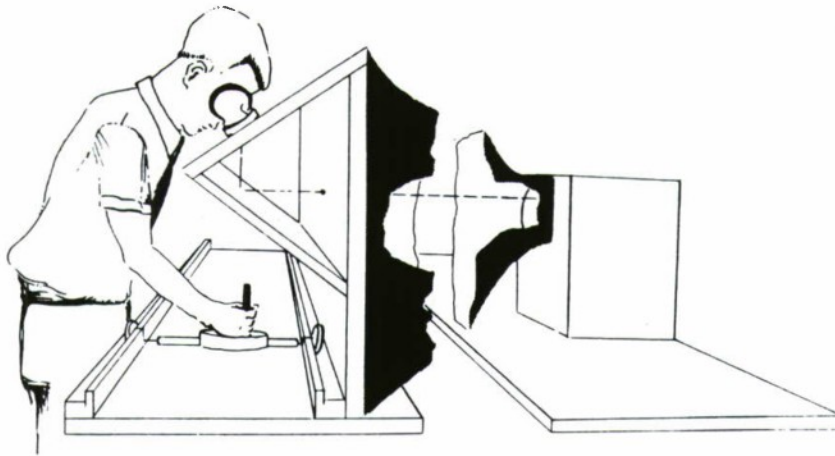
Figure 4.– Experimental setup for studying adaptation to visual displacement and delay.
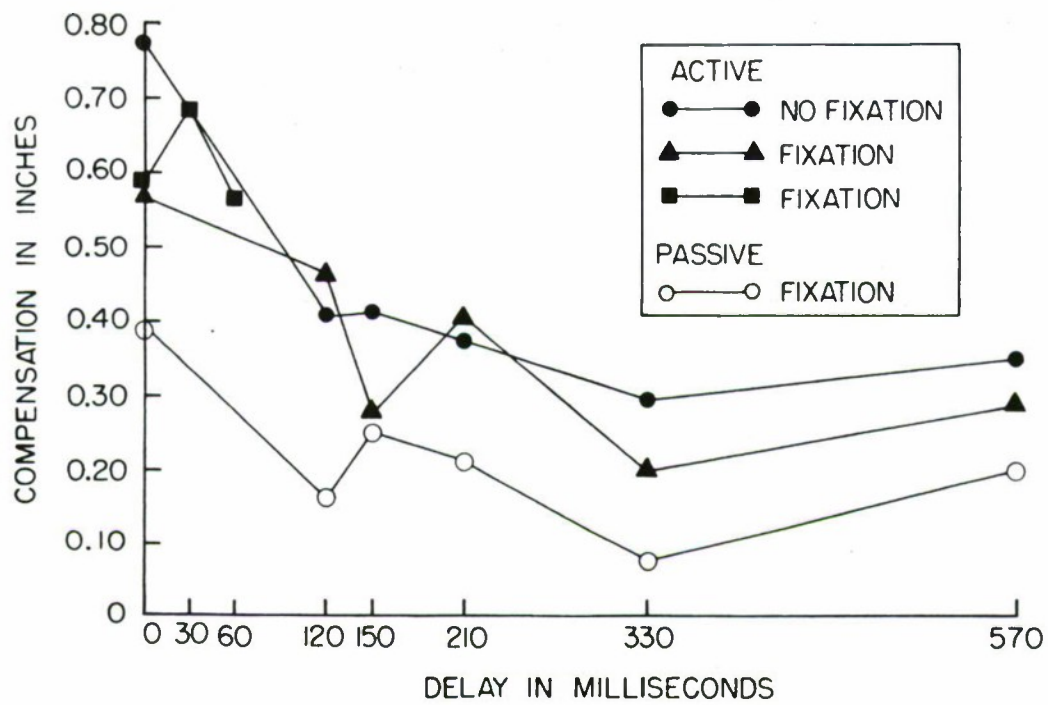


Figure 5.– Results of experiments.

# ADAPTING TO VARIABLE PRISMATIC DISPLACEMENT[1]

Robert B. Welch and Malcolm M. Cohen
NASA Ames Research Center
Moffett Field, California

## SUMMARY

In each of two studies subjects were exposed to a continuously changing prismatic displacement with a mean value of 19 prism diopters ("variable displacement") and to a fixed 19-diopter displacement ("fixed displacement"). In Experiment 1, we found significant adaptation (post-pre shifts in hand-eye coordination) for fixed, but not for variable, displacement. Experiment 2 demonstrated that adaptation can be obtained for variable displacement, but that it is very fragile and will be lost if the measures of adaptation are preceded by even a very brief exposure of the hand to normal or near-normal vision. Contrary to the results of some previous studies, we did not observe an increase in within-S dispersion of target-pointing responses as a result of exposure to variable displacement.

## INTRODUCTION

Human observers who are allowed to view their actively moving hands through an optical medium that displaces, inverts, right-left reverses, or otherwise rearranges the visual field reveal significant adaptive changes in hand-eye coordination (Welch, 1978). For example, the initial errors made when one looks through a wedge prism and attempts to touch a target are typically corrected in a matter of minutes. Depending on the nature of the exposure conditions, this prism-adaptive shift in hand-eye coordination can be based on changes in (1) the felt position of the limb (e.g., Harris, 1965); (2) visual localization (e.g., Craske, 1967); or (3) the algebraic sum of both of these events (e.g., Wilkinson, 1971).

An alternative to prismatic displacement of constant strength (which may be referred to as "fixed displacement") is one that varies continuously in both magnitude and direction ("variable displacement"). It has been shown by Cohen and Held (1960) that active exposure to a variable displacement in the lateral dimension with a mean value of zero fails to produce an adaptive shift in the *average* location of the subject's repeated target-pointing attempts, although it does appear to increase the *variability* of these responses around the mean. The latter observation has been interpreted as a degradation in the precision of hand-eye coordination.

The absence of adaptation to this form of variable displacement should not come as a surprise, since, over the course of the prism exposure period, there is no net prismatic displacement to which one can adapt. What remains to be determined, however, is whether it is possible to adapt to a situation of variable displacement in which the mean value is significantly different from zero, since in this case it is at least plausible for such adaptation to occur. The aim of the present

investigation was to answer this question and, in addition, to compare the magnitude of such adaptation with that produced by comparable fixed prismatic displacement.

## METHOD

### General Design

Two experiments were carried out. In both, subjects were used as their own control under conditions of fixed and variable prism exposure to the same average displacement (19 prism diopters). This comparison is seen in figure 1. Experiment 1 also included the between-group factor of direction (up vs. down) of the optical displacement of the hand that was present during exposure. Prism adaptation was indexed by the difference between pre- and postexposure target-pointing accuracy without visual feedback (visual open-loop).[2] Also obtained were post-pre differences in the within-S variability (standard deviation) of target-pointing over the 10 pre- and 10 postexposure trials. Finally, potential intermanual transfer of the prism-adaptive shifts in target-pointing was examined by testing both exposed and nonexposed hands.

### General Procedure and Apparatus

At the outset of the testing period, subjects sat at a table with faces pressed into the frame of a pair of prismless (normal-vision) goggles built into a box. Looking into this box, they viewed the reflection of a back-illuminated 1- by 1-in. cross, the apparent position of which was straight ahead at approximately eye level and at a distance of 48 cm, nearly identical to that of a vertically positioned 12- by 12-in. touch pad. For the preexposure (and later the postexposure) measures of target-pointing accuracy, subjects pointed alternately with the right and left index fingers (10 responses each), attempting to contact the touch pad at a place coincident with the apparent center of the cross. The inter-response interval was approximately 3 sec. The mirror blocked the view of the pointing hand, thereby precluding error-corrective visual feedback. When subjects touched the pad, the X and Y coordinates of the finger's position were immediately signaled and written to a floppy disk, using a program supported by an Apple II Plus computer.

During the prism-exposure period, the prismless goggles were replaced by binocular prisms (variable or fixed) and the mirror was moved out of the way, allowing subjects to see the touch pad as well as the hand when it was brought into view. In addition, a hand-movement guide consisting of a vertical rod was situated parallel to and approximately 9 cm away from the surface of the pad.

The exposure period consisted of a series of 55-sec cycles. During the first half of each cycle, subjects, who were looking through the (upward- or downward-displacing) prisms, actively moved the preferred hand up and down along the rod, fixating the limb at all times. They grasped the rod with the thumb hooked around the rod and the palm of the hand facing them. Hand

---

[2]An attempt was also made to obtain measures of prism-adaptive shifts in felt-limb position. During the pre- and postexposure periods, subjects (with eyes shut) were to try to place the right and left index finger (alternately) at a position on the touch pad that they felt to be directly in a horizontal line with an imaginary point in the center of the bridge of their nose. Unfortunately, many subjects reported that they approached this task as if it were merely another form of target-pointing. Furthermore, their responses were erratic and the data were difficult to interpret. For these reasons, the results from these measures have been omitted from this report.

movements were made to the beat of a 1-Hz electronic metronome; the limb was moved up on the first beat, down on the next beat, and so forth, for exactly 27.5 sec. Then for the next 27.5 sec the subjects rested the hand on the table and fixated the cross while looking through the goggles, which were now set to produce displacement in the opposite direction. This was followed by 27.5 sec of observed hand movement, with the direction of prismatic displacement returned to its original state. Subjects alternated between these two displacements for a total of nineteen 55-sec cycles (17:25 min). Finally, postexposure measures of target-pointing accuracy were obtained in the same manner as the preexposure measures.

The conditions of fixed downward and fixed upward displacement were achieved by means of paired base-up and base-down wedge prisms, respectively. The prisms were attached to a sliding panel that moved them to a position directly in front of the goggle eyepieces. Variable displacement in the vertical dimension was produced by a pair of binocular, motor-driven Risley prisms which rotated in opposite directions; the net result was a binocular optical displacement that continuously changed in the vertical dimension over a range of ±30 diopters (±17.1°).

Measures of potential prism-adaptive shifts in target-pointing accuracy in the vertical dimension were obtained by subtracting (for each hand separately) the mean of the 10 preexposure responses from the mean of the 10 postexposure responses. Potential prism-induced changes in within-S variability of target pointing were determined by subtracting the standard deviation of a given subject's 10 preexposure measures (for a particular hand) from the standard deviation of the corresponding 10 postexposure measures.

## EXPERIMENT 1

### Design

Twelve subjects (8 males and 4 females, ages 19-33) were randomly divided into two 6-subject groups. For one group the visual field was displaced upward during that half of each cycle in which the subject viewed the actively moving hand; for the other, the field was displaced downward. Subjects were tested individually in two conditions–variable displacement and fixed displacement–occurring 48 hr apart. The order of the two conditions was counterbalanced across subjects.

### Procedure

Following the preexposure measures of open-loop target pointing, the mirror was removed and subjects looked through prismless (i.e., nondisplacing) goggles while undergoing the nineteen 55-sec cycles. On each cycle the hand was viewed for 27.5 sec, followed by 27.5 sec of viewing the target cross while the hand was resting on the table out of view. The purpose of this long period of normal vision was to establish an accurate and reliable baseline measure of each subject's perception of the hand's location under nondistorted visual circumstances before introducing the prismatic displacement. After a short rest break, subjects repeated the procedure, but this time they viewed the moving hand through prisms that were set either for fixed or for variable displacement. In order to reduce the possibility of significant loss of adaptation through spontaneous decay, the postexposure measures were obtained immediately after the subjects had viewed the prismatically

displaced hand, which necessitated terminating the last cycle after the first 27.5 sec. It is important to note that because of this procedural decision the last view of the hand for the fixed displacement condition was one of 19 diopters' displacement, while for the variable-displacement condition it entailed little or no displacement (see fig. 1).

## Results

As shown in figure 2, prism-adaptive shifts in target-pointing accuracy for the exposed hand were obtained in the fixed, but not in the variable, displacement condition for both the upward and downward displacement groups. The finding of adaptive post-pre shifts for both directions of displacement confirms that these changes represent adaptation to the prisms *per se,* rather than some form of "drift" of pointing accuracy over time due to fatigue or other factors unrelated to the prismatic displacement. Analysis of variance revealed main effects for Direction (up/down), $F(1,4) = 14.49$, $p = 0.22$, and Displacement (variable/fixed), $F(1,4) = 30.01$, $p < 0.01$, and for the Direction/Displacement interaction, $F(1,4) = 82.14$, $p < 0.001$. Figure 2 indicates that the difference between the variable and fixed displacement conditions was greater for the upward displacement group. There was no main effect for order, nor was this factor involved in any interactions. Adaptation for the nonexposed hand (due to intermanual transfer) was obtained only for the fixed/upward displacement condition.

No statistically significant post-pre shifts in the dispersion (standard deviations) of target pointing were obtained for either hand in any condition.

Finally, for none of the conditions was there evidence of any decay of adaptation over the 10 postexposure trials for either hand.

## Discussion

Since adaptation occurred for fixed but not variable displacement, the answer to the original experimental question would seem to be that human observers are *not* capable of adapting to nonzero variable displacement, at least with exposure periods of the length used here. There is, however, an alternative possibility, based on the fact that for subjects in the variable-displacement condition, the last experience during the prism exposure period was of normal or near-normal vision (fig. 1). It may be suggested that the adaptation produced in this experiment (or perhaps specifically in the variable-displacement condition) is quite *fragile* and therefore easily destroyed by subsequent exposure to normal vision. If so, then one could suppose that adaptation was actually produced in both conditions, but eliminated for the variable-displacement condition because of the "unlearning" that occurred at the very end of the exposure period. Experiment 2 attempted to examine this possibility by asking the following question: Does the difference in adaptation in favor of fixed displacement that was obtained in Experiment 1 remain when the exposure period for the variable-displacement condition is caused to end on maximum displacement, rather than on no displacement?

# EXPERIMENT 2

## Design

Six subjects (2 males and 4 females, ages 21-39) were used as their own control in conditions of variable and fixed displacement in the upward direction only. The two conditions were separated by 48 hr and their order of occurrence counterbalanced across subjects.

## Procedure

During the prism-exposure period, subjects viewed the preferred hand in the same manner as in Experiment 1, with the addition of one extra half-cycle. The latter ended after only 13.75 sec, which meant that the prismatic displacement for the variable condition was at its maximum of 30 diopters while the displacement for the fixed condition remained at its constant level of 19 diopters (see fig. 1).

Pre- and postexposure measures of target-pointing accuracy for both hands were taken in the same manner as in Experiment 1.

## Results

As may be seen in figure 3, prism-adaptive shifts in target-pointing accuracy were found for both variable and fixed-displacement conditions and both exposed and nonexposed hands. All of the post-pre shifts were significantly different from zero, but there were no main effects for the factors of Hand (exposed/non-exposed) or Displacement (variable/fixed), nor any interactions. Once again, no prism-induced changes in target-pointing precision (within-S standard deviations) or postexposure decay of adaptation were observed.

## Discussion

The results of Experiment 2 are consistent with the "fragility hypothesis," since when the most recent visual experience in the variable-displacement condition was of maximum displacement, adaptation was substantial and, indeed, as great as that produced by fixed displacement. An interesting secondary finding was the large amount (i.e., 100%) of intermanual transfer produced.

# CONCLUSIONS

The present study has demonstrated that human subjects are capable of adapting their hand-eye coordination to nonzero variable displacement, although this adaptation is quite easily destroyed. It is possible, of course, that this fragility is unique to the current situation in which the prism-exposure task did not involve visual error-corrective feedback and exposure periods were repeatedly interrupted by rest periods. Furthermore, the present design does not allow us to

exclude the possibility that the adaptation produced in the fixed-displacement condition was also fragile and would therefore have been quickly eliminated by exposure to normal vision.

A surprisingly large amount of adaptation was observed for the nonexposed hand, especially in Experiment 2. This may have been due to the use of alternating exposure and rest periods, since "distribution of practice" has been demonstrated to facilitate intermanual transfer of prism adaptation (e.g., Cohen, 1973). Such intermanual transfer has frequently been used as evidence that prism-adaptive changes in *vision* have occurred. Evidence against this interpretation of the present observations, however, comes from studies (e.g., Uhlarik and Canon, 1971) showing that prism exposure not involving target-pointing, as in this experiment, is generally ineffective in producing this kind of adaptation. An alternative interpretation of intermanual transfer of prism adaptation is that it represents a central change in motor programming that is usable, at least to some extent, by the nonexposed hand.

Contrary to the results of Cohen and Held (1960), neither of the present experiments revealed an increase in the dispersion of target pointing as a result of exposure to variable displacement. Two explanations for this failure to replicate may be proposed. First, it is possible that the presence of only one target for the pre- and postexposure trials (in contrast to the four used by Cohen and Held, 1960) was conducive to a "stereotyping" of target-pointing responses. Such a potential constraint on trial-to-trial variability would be likely to counteract any disruptive effects that variable displacement might have on the within-subject dispersion of responses. Second, the present exposure period was relatively brief in comparison to that used in the Cohen-Held experiment. Indeed, in the latter, no increase in dispersion was obtained until after 30 min of variable displacement. In the present experiment, actual exposure to the hand (excluding the 27.5-sec "rest" periods) amounted to only a little over 8 min.

It is of interest to speculate why variable prismatic displacement should produce adaptation that is so easily destroyed (assuming that future research supports this conclusion). One possibility is that exposure to variable displacement causes the adaptive system to be quite labile and therefore easily changed, even by very brief exposures to new visual displacements or to normal vision. This interpretation fits with the finding by Cohen and Held (1960) of degraded hand-eye precision after exposure to variable displacement, but is weakened by the present failure to replicate the Cohen-Held observation.

A second possibility is that subjects exposed to variable-displacement experience only "visual capture," a nearly instantaneous shift in felt-limb position when viewing the prismatically displaced hand (Welch and Warren, 1980). Since visual capture is extremely fragile, it will be destroyed by even a brief exposure to normal vision and will also rapidly decay when view of the hand is precluded. The quick decay of visual capture, however, contrasts with the absence of postexposure decay in either of the present experiments, rendering this interpretation questionable.

The most likely explanation of the present results is that when human observers are actively exposed to a systematically changing prismatic displacement, they acquire the ability to adapt (or readapt) nearly instantaneously, as required. Such presumptive adaptive flexibility would represent a clear advance over the situation with fixed displacement, since the latter involves relatively slow acquisition of adaptation and the presence of substantial aftereffects upon return to normal vision. In short, it is possible that prolonged exposure to variable displacement provides the observer with the ability to shift from one set of visuomotor relationships to another with a minimum of disruption. An experiment to evaluate this interpretation is currently being implemented.

# REFERENCES

Cohen, M. M. (1973). Visual feedback, distribution of practice, and intermanual transfer of prism aftereffects. Percept. Motor Skills, 37, 599-609.

Cohen, M. M., and Held, R. (1960). Degrading visual-motor coordination by exposure to disordered re-afferent stimulation. Paper presented at meetings of Eastern Psychological Association, New York City.

Craske, B. (1967). Adaptation to prisms: Change in internally registered eye-position. British J. Psychol. 58, 329-335.

Harris, C. S. (1965). Perceptual adaptation to inverted, reversed, and displaced vision. Psychol. Rev., 72, 419-444.

Uhlarik, J. J., and Canon, L. K. (1971). Influence of concurrent and terminal exposure conditions on the nature of perceptual adaptation. J. Exp. Psychol., 91, 233-239.

Welch, R. B. (1978). Perceptual modification: Adapting to altered sensory environments. New York: Academic Press.

Welch, R. B., and Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. Psychol. Bull., 88, 638-667.

Wilkinson, D. A. (1971). Visual-motor control loop: A linear system? J. Exp. Psychol., 89, 250-257.
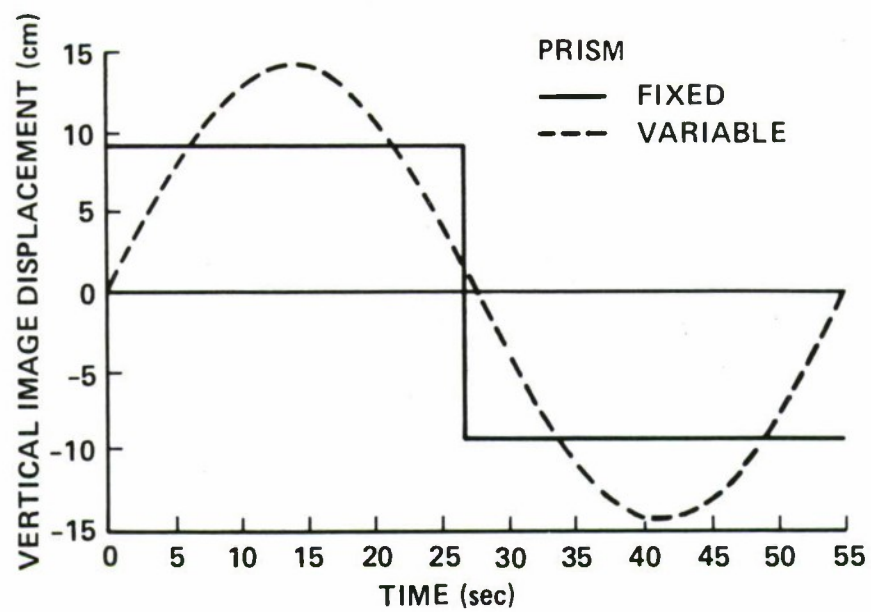
Figure 1.– Prismatic exposure conditions: Fixed and variable prism displacements.
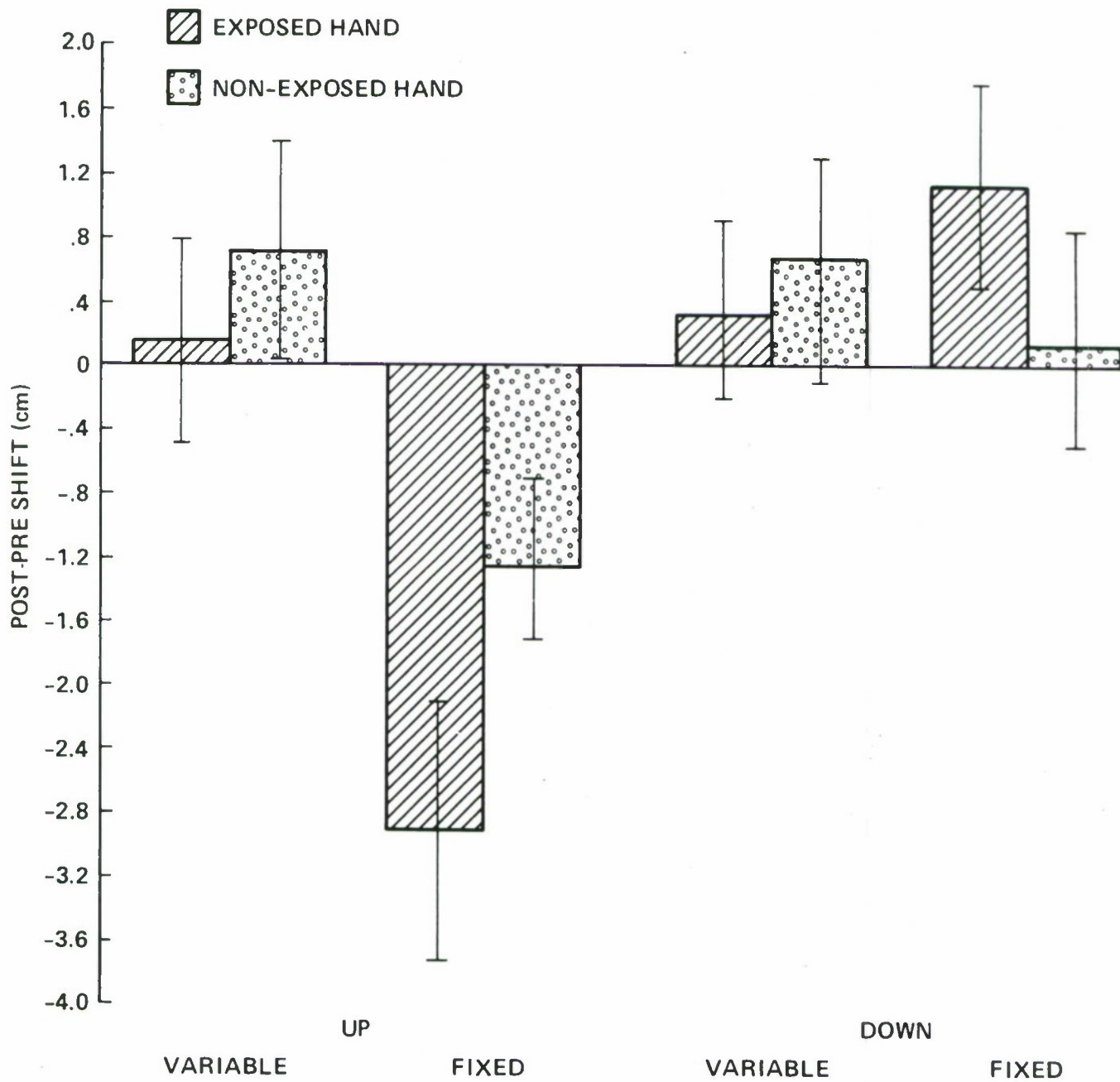
Figure 2.– Experiment 1: Post-pre shifts (cm) in target-pointing accuracy.
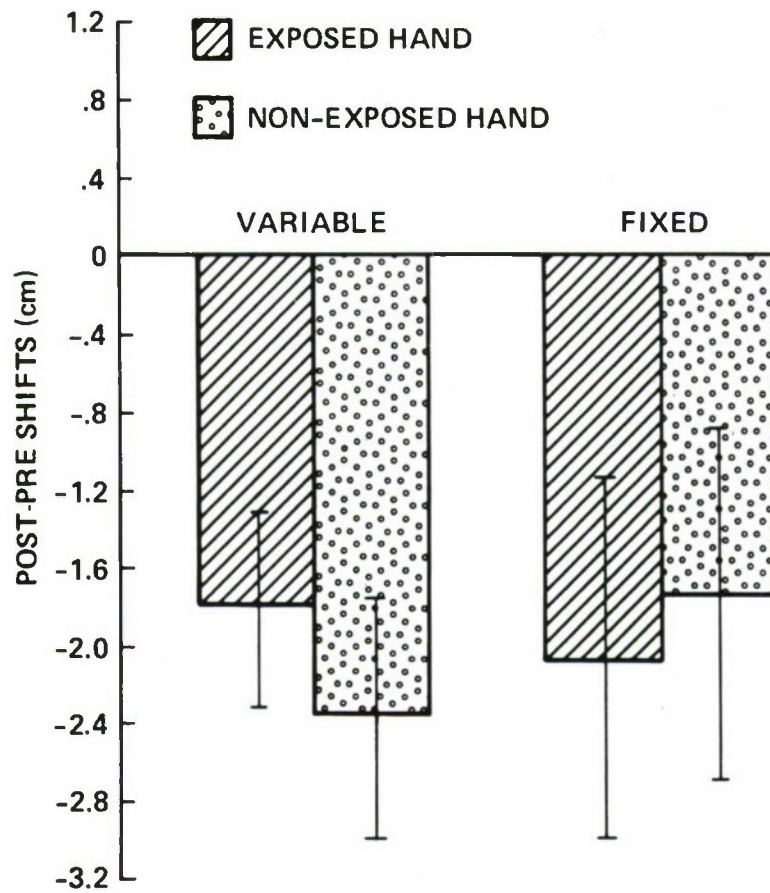
Figure 3.– Experiment 2:  Post-pre shifts (cm) in target-pointing accuracy.

# VISUAL ENHANCEMENTS IN PICK-AND-PLACE TASKS: HUMAN OPERATORS CONTROLLING A SIMULATED CYLINDRICAL MANIPULATOR

Won S. Kim, Frank Tendick, and Lawrence Stark
University of California
Berkeley, California

## ABSTRACT

A visual display system serves as an important human/machine interface for efficient tele-operations. However, careful consideration is necessary to display three-dimensional information on a two-dimensional screen effectively. A teleoperation simulator is constructed with a vector-display system, joysticks, and a simulated cylindrical manipulator in order to evaluate various display conditions quantitatively. Pick-and-place tasks are performed, and mean completion times are used as a performance measure. Two experiments are performed. First, effects of variation of perspective parameters on a human operator's pick-and-place performance with monoscopic perspective display are investigated. Then, visual enhancements of monoscopic perspective display by adding a grid and reference lines are investigated and compared with visual enhancements of stereoscopic display. The results indicate that stereoscopic display does generally permit superior pick-and-place performance, while monoscopic display can allow equivalent performance when it is defined with appropriate perspective parameter values and provided with adequate visual enhancements. Mean-completion-time results of pick-and-place experiments for various display conditions shown in this paper are observed to be quite similar to normalized root-mean-square error results of manual tracking experiments reported previously.

## INTRODUCTION

Visual display systems serve as an important human/machine interface for efficient teleoperations in space, underwater, and in radioactive environments.[1-4] Closed-circuit television systems, presenting two-dimensional (2-D) images captured by remote video cameras, have been commonly used for these visual displays. As technology evolves from manually controlled teleoperations to sensor/computer-aided advanced teleoperations[5,6] or telerobotics,[7-11] graphics displays have been drawing attention as a means to provide an enhanced human/machine interface. A graphic display can present an abstract portrayal of the working environment or state of the control system based on sensor signals and a data base.[2,12] A force-torque display[13] and a "smart" display[14] are examples of graphic displays developed for efficient teleoperations.

There are two types of visual displays: monoscopic and stereoscopic. The stereoscopic display provides two slightly different perspective views for the human operator's right and left eyes. A stereoscopic view enables the human to perceive depth by providing a distinct binocular depth cue called stereo disparity. Some earlier studies with television displays showed that stereoscopic displays, as compared to monoscopic displays, did not provide significant advantage in performing some telemanipulation tasks.[15-17] Careful recent studies,[18,19] however, indicated that stereo performance was superior to mono under most conditions tested, while the amount of improvement

varied with visibility, task, and learning factors. These results showed that the advantage of the stereoscopic television display became pronounced with increased scene complexity and decreased object visibility.

Monoscopic and stereoscopic graphic displays were recently compared by employing three-axis manual tracking tasks.[20,21] Root-mean-square (rms) tracking error was used as a performance measure for quantitative evaluation. Results were consistent with previous television display results, indicating that stereoscopic graphic displays did generally permit superior tracking performance, while monoscopic displays allowed equivalent performance when they were defined with appropriate perspective parameters and provided with adequate visual-enhancement depth cues such as reference lines.

The purpose of our present study is to examine generality or consistency of the above results. A three-axis pick-and-place task, instead of the three-axis manual tracking task, is employed in our present study as a realistic teleoperations task. Two experiments similar to those in reference 21 are performed. In the first experiment, we quantitatively evaluate monoscopic perspective display by investigating individual effects of perspective parameters. Perspective projection alone, however, does not provide sufficient three-dimensional (3-D) depth information for monoscopic display. Thus, a 5-line-by-5-line horizontal grid representing a base plane and a vertical reference line representing vertical separation from the base plane are introduced as two visual-enhancement depth cues. In the second experiment, we investigate effects of these two visual-enhancement depth cues on pick-and-place performance for both monoscopic and stereoscopic displays.

## METHODS

In order to evaluate various display conditions quantitatively, a teleoperations simulator is constructed with a vector-display system, joysticks, and a simulated cylindrical manipulator. Figure 1 shows a schematic diagram of the experimental setup, with which three-axis pick-and-place tasks are performed.

### Real-Time Simulation of The Manipulator

The Hewlett-Packard 1345A vector-display module is used for real-time dynamic display. It has high resolution (2048 x 2048 addressable data points), and high vector-drawing speed (8194 cm of vectors at 60-Hz refresh rate). It also has a fast vector-updating speed (approximately 10 $\mu$sec/vector), communicating with a host computer through a 16-bit parallel I/O port. Two isotonic (displacement) joysticks are employed for the Cartesian position control of the manipulator gripper. An LSI-11/23 computer with the RT-11 operating system is used as a host computer. It performs computations for the simulated manipulator motion and perspective or stereoscopic display, and measures task completion time.

The human operator indicates the desired gripper position of the manipulator in robot base Cartesian coordinates by using three axes of the two joysticks. The computer senses the joystick displacements through 12-bit A/D converters. The joystick gain for each axis is chosen to be 1 so that the full range of the joystick displacement for each axis corresponds to the full movement range of the gripper position for the corresponding axis. The computer transforms the desired

gripper position in Cartesian coordinates to the desired joint angle ($\theta_1$ for the revolute joint 1) and joint slidings (d2 and d3 for the prismatic joints 2 and 3) by employing the inverse kinematic position transformation. The next two sections describe how to present 3-D information of the manipulator on the 2-D display screen.

## Monoscopic Perspective Display

A monoscopic perspective display can be constructed by a perspective projection of an object onto the view plane (projection plane) followed by a mapping of the view plane onto the screen.[22] There are two approaches to obtaining the perspective projection of an object. One is to leave the object stationary and choose a desired viewpoint and a projection plane, called the viewpoint-transformation method. The other approach is to fix the viewpoint and transform the object, called the object-transformation method. These two approaches are mathematically equivalent.[21,23] The latter will be described here.

In order to derive the perspective display formulas based on the object-transformation method, a right-handed XYZ world coordinate system is established. The viewpoint is fixed at the origin $(0, 0, 0)$ and the view plane at the $z = -d$ plane. Perspective projection can be obtained by three transforms: rotation R, translation T, and perspective transform P.

Initially, an object is located so the view reference point of the object is at the origin. Then the object is appropriately rotated and translated to achieve the desired viewing angles and distance. In general, an arbitrary orientation of an object can be described by successive principal-axis rotations about the Y, X, and Z axes.

$$R = Rot(Y, -\theta_1) \; Rot(X, \theta_2) \; Rot(Z, \theta_3) \tag{1}$$

where the yaw, pitch and roll angles are $-\theta_1, -\theta_2$, and $\theta_3$, respectively. It can be shown that the yaw and pitch angles used in the object transformation approach are equivalent to the azimuth and elevation angles in the viewpoint-transformation approach.[21]

For simplicity, 4-space homogeneous coordinate transformations are used. The rotation of a point at position $(x, y, z)$ to a new position $(x', y', z')$ can be described by

$$(x', y', z', 1) = (x, y, z, 1) \, R \tag{2}$$

where

$$R = \begin{bmatrix} R_{11} & R_{12} & R_{13} & 0 \\ R_{21} & R_{22} & R_{23} & 0 \\ R_{31} & R_{32} & R_{33} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3}$$

From equation (1), each element of the 4 x 4 matrix R can be calculated as $R_{11} = C_1C_3 - S_1S_2S_3$, $R_{12} = -C_1S_3 - S_1S_2C_3$, $R_{13} = S_1C_2$, $R_{21} = C_2S_3$, $R_{22} = C_2C_3$, $R_{23} = S_2$, $R_{31} = -S_1C_3 - C_1S_2S_3$, $R_{32} = S_1S_3 - C_1S_2C_3$, $R_{33} = C_1C_2$. $S_i$ and $C_i$ denote $\sin \theta_i$ and $\cos \theta_i$, respectively.

After the rotation, the object is translated by D along the negative Z axis.

$$T = \text{Trans}(0, 0, -D) \tag{4}$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -D & 1 \end{bmatrix}. \tag{5}$$

The length D represents the distance from the viewpoint to the view reference point, called the object distance.

The UV coordinate system is embedded in the view plane. Perspective transformation of a point Q (x, y, z) in the world coordinate to its projection $Q_p$ (u, v) on the view plane can be described by

$$(x', y', z', w) = (x, y, z, 1) P \tag{6}$$

$$(u, v) = (x'/w, y'/w) \tag{7}$$

where

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1/d \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{8}$$

The symbol d denotes the view plane distance from the viewpoint. Increase of the view plane distance results in uniform magnification of the perspective projection. Thus, d can be specified in terms of the zoom or magnification factor, which can be defined as $M = d/D$. Distance d can also be specified in terms of field-of-view (fov) angle, which is the angle at the viewpoint subtended by the view-plane window. If the view plane window is specified as a square region $(u_{min}, u_{max}, v_{min}, v_{max}) = (-1, 1, -1, 1)$, then the fov angle is related to the view-plane distance by $d = \cot(\text{fov}/2)$. The perspective projection obtained with a wide fov angle is similar to the picture taken by a wide-angle camera lens, and a narrow fov angle is similar to one taken by a tele-photo lens.

After the object is projected onto the view plane, mapping of the view plane onto the physical display screen is performed. Mapping of a point from (u, v) in the UV coordinate to $(x_s, y_s)$ in the screen coordinate can be achieved by appropriate translations and scalings:

$$x_s = \text{VSX } u + \text{VCX} \tag{9}$$

$$y_s = \text{VSY } v + \text{VCY} \tag{10}$$

where VSX and VSY are scaling factors, and VCX and VCY are translation factors.

# Stereoscopic Display

The monoscopic display does not give true depth perception. The human brain merely interprets the 2-D monoscopic picture as 3-D space. The stereoscopic display presents two views of an object on the display: one for the right eye, and the other for the left. This pair of pictures is called a stereo pair or a stereogram. The human operator views a stereogram through a stereoscope.[24] Most people can fuse the stereo pair into one 3-D image, perceiving relative depth by the human stereoscopic vision ability. The stereoscope is composed of two converging lenses and a supporting frame (septum) separating right and left views. As illustrated in figure 2, two converging lenses form the image of the stereo pair onto the image plane behind the actual display screen, which can provide fairly correct accommodation and convergence conditions for the human eyes, if the geometrical and optical conditions are appropriately arranged.

In order to obtain the formulas for the stereoscopic display, an XYZ coordinate system is established with its origin in the middle of the two optical centers for the right and left eyes, as depicted in figure 2. The display screen, on which a stereogram is presented, is located at the picture plane (view plane, projection plane) $z = -d$. The two converging lenses of the stereoscope form the virtual image of the stereogram on the image plane $z = -D$. By denoting the focal length of the binocular lens as $F$, the converging lens formula yields

$$\frac{1}{d} - \frac{1}{D} = \frac{1}{F} \tag{11}$$

When $D$ is infinity, $d = F$. When $D = 40$ cm and $F = 20$ cm, $d = 13.3$ cm.

As in the object-transformation approach used previously for the monoscopic perspective display, the object is initially located so the view-reference point of the object is at the origin. Then the object is appropriately rotated and translated using equations (3) and (5) to achieve the desired viewing angles and distance.

Denoting the interocular distance (IOD) (approximately 5.5 to 6.5 cm), we can express the positions of the two optical centers by $(x_{or}, 0, 0)$ for the right eye and $(x_{ol}, 0, 0)$ for the left eye, where $x_{or} = IOD/2$, and $x_{ol} = -IOD/2$. The projection of a point $P(x, y, z)$ onto the view plane for each eye is formed at the intersection of the projection line with the view plane. By representing the right and left projection points by $P_r(x_r, y_r)$ and $P_l(x_l, y_l)$, respectively, the following equations can be obtained:

$$x_r = x_{or} + (x - x_{or})(-d/z) \tag{12}$$

$$x_l = x_{ol} + (x - x_{ol})(-d/z) \tag{13}$$

$$y_r = Y_1 = y(-d/z) \tag{14}$$

Finally, these projection points on the projection plane can be mapped onto the physical screen coordinates by appropriate translations and scalings.

# Experimental Procedures

Two sets of experiments were performed, varying perspective parameters and visual enhancement conditions. In both experiments, subjects were seated in front of the display (on which the manipulator, the objects to pick up, and the boxes to place them in were presented) (fig. 3), and the subjects were asked to perform three-axis pick-and-place tasks. The subjects controlled the manipulator using two joysticks to pick up each object with the manipulator gripper and place it in the corresponding box. One hand, using two axes (forward-backward and right-left) of one joystick, controlled the gripper position for the two axes parallel to the horizontal base plane. The other hand, using one axis (forward-backward) of the other joystick, controlled the vertical axis.

Each of the four objects (point targets A, B, C, D) was positioned randomly within the manipulator reach space. Each object position was marked by a tiny diamond and a letter. Picking up an object was accomplished when the manipulator gripper touched the object within the boundary of the error tolerance, defined by a hypothetical cube. The size of the cube was set so that the picking process was neither too easy nor too hard within the range of experimental variation. Accomplishment of picking up an object was indicated by doubling the object letter. Thereafter, the object moved together with the gripper until it was placed in the right box. Placing an object was accomplished by touching the correct box with the gripper, similar to the picking process. After the touch, the object symbol letter became single again, and the object remained in the box, while the gripper was free to move for the next operation.

One run of the pick-and-place task consisted of five sessions of four pick-and-place operations in order from object A to D, totaling 20 pick-and-place operations.

Perspective Parameter Experiment. In this experiment, we investigated the effects of different perspective parameters on the human operator's pick-and-place performance with monoscopic perspective display. The five perspective parameters, azimuth, elevation, roll, fov angle, and object distance were independently varied, keeping the other variables fixed at their nominal values. The nominal perspective parameter values were chosen as elevation = $-45°$, azimuth = $0°$, roll = $0°$, fov angle = $12°$, and object distance = 40 cm.

Experimental variables were varied as follows: (1) seven elevation angles: $0°$, $-15°$, $-30°$, $-45°$, $-60°$, $-75°$, and $-90°$; (2) eight azimuth angles: $-135°$, $-90°$, $-45°$, $0°$, $45°$, $90°$, $135°$, and $180°$; (3) eight roll angles: $-135°$, $-90°$, $-45°$, $0°$, $45°$, $90°$, $135°$, and $180°$; (4) five fov angles: $8°$, $12°$, $24°$, $48°$, and $64°$, (5) four object distances: 30, 40, 80, and 160 cm.

The monoscopic perspective presentation with the nominal perspective parameters is shown in figure 3. Some examples of variations in perspective parameter values used in this experiment are shown in figure 4. In this experiment, a 5-line-by-5-line horizontal grid and vertical reference lines were always presented. The experiment was run with each of the 32 experimental conditions presented in random order. There were two runs of 20 pick-and-place operations per condition for each subject. For the monoscopic conditions, the subjects were seated 40 cm in front of the display screen.

Visual Enhancement Experiment. In this experiment, effects of visual enhancements on the human operator's pick-and-place performance were investigated. The visual-enhancement depth cues used for both monoscopic and stereoscopic displays were a grid and reference lines. Three-axis pick-and-place tasks were performed for four visual-enhancement conditions at each of five different perspective parameter conditions with both monoscopic and stereoscopic displays. The

four visual-enhancement conditions were: GL (presence of both grid and reference line), L (reference line only), G (grid only), and O (neither). The five perspective parameter conditions used were: (1) 0° in elevation, (2) –90° in elevation, (3) nominal perspective parameter values, (4) 45° in azimuth, and (5) 80 cm in object distance.

Monoscopic presentations for the four visual-enhancement conditions with the nominal perspective parameters (condition III) are shown in figure 5. Monoscopic presentations for the five perspective parameter conditions, when both grid and reference lines are presented (condition GL), are shown above the mean completion time plot in figure 8. A stereoscopic presentation with the nominal perspective parameters, when both grid and reference lines are presented, is shown in figure 6. The experiment was run first with each of the 20 monoscopic display conditions presented in random order, then with each of the 20 stereoscopic display conditions presented in random order. There were two runs of 20 pick-and-place operations per condition for each subject.

In the monoscopic display conditions, subjects were seated 40 cm in front of the screen. In the stereoscopic display conditions, subjects were seated 13.3 cm in front of the screen, viewing the stereogram through the stereoscope. The focal length of the converging lens of the stereoscope was 20 cm, and thus the virtual image of the stereogram was formed at 40 cm from the lens (by eq. (11)).

## Subjects

Two young adult male subjects with normal stereo vision participated in each of the two experiments. Each subject was trained for at least 5 hr before the experiments to saturate the "learning" effect. During the training period, mean completion times were regularly checked to see whether the subject reached an asymptotic, steady-state, pick-and-place performance. However, during the actual experiment, mean completion times were not checked until all the experimental runs were completed. Each subject repeated the experiment once more in order to examine intra-subject variation as well as inter-subject variation.

## EXPERIMENTAL RESULTS

Mean completion time was used as the performance measure in our pick-and-place tasks. Each of the mean completion time data points in figures 7 and 8 is the average obtained from one run of 20 pick-and-place operations.

The experimental results for two subjects with two runs each plotted in figure 7 with mean completion time as the ordinate and perspective parameter values as the abscissa. The effects of elevation, azimuth, roll, fov angle, and object distance are plotted in figure 7 (a), (b), (c), (d), and (e), respectively.

The experimental results for two subjects with two runs each are shown in figure 8. Mean completion time (ordinate) is plotted for the various display conditions (abscissa). The monoscopic display data are marked by squares and dashed lines, and the stereoscopic display data are marked by filled diamonds and solid lines. The five separate columns represent five different per-

spective parameter settings, conditions 1-5. Each column has four different visual-enhancement conditions, GL, L, G, and O.

## DISCUSSION

### Effects of Perspective Parameters

The mean-completion-time plots of figure 7 show the effects of variation of perspective parameters on pick-and-place performance. Plot (a) shows that as the elevation angle approaches 0° or –90°, mean completion time increases. This is due to the loss of one axis' position information. Performance at –90° elevation was better than at the 0° extreme because the perspective view at –90° elevation made it possible to see some of the height of the reference line if it was not near the center of the projected image. Thus, there was a partial view of the "lost" axis. Plot (b) shows that as the azimuth angle exceeds the range of –45° to +45°, the mean completion time increases markedly. An azimuth angle other than 0° implies rotation of the display reference frame relative to the joystick control axes, thus making the joystick control more difficult compared to the 0° azimuth angle. When the azimuth angle is beyond –45° to +45°, it is difficult for the human operator to compensate. Performance is especially poor when the azimuth angle is about –90° or +90°, even worse than the case when azimuth angle is 180°. At 180° azimuth angle, the human operator uses inversion rather than rotation. Plot (c) shows that change in roll angle produces an effect similar to changing the azimuth angle, because of analogous disorientation. Plots (d) and (e) show that as the fov angle or the object distance increases, and the displayed object picture becomes smaller, task performance degrades.

### Effects of Visual Enhancements

The results of the visual-enhancement experiment appear in figure 8 (a) and (b). Monoscopic display results in columns I and II show that when the elevation angle is 0° or –90°, the mean completion times are very long, even with grid or reference line enhancements. This is because position information for one axis is lacking, and the subject must sweep the gripper along that axis until it touches the correct position. At –90° elevation, the reference lines almost disappear. At 0° elevation, the grid appears as a single line. Monoscopic display results in columns III, IV, and V show that by choosing adequate elevation angles, mean completion times can be shortened, and fast pick-and-place performance can be attained with monoscopic perspective display, if reference lines are provided (GL, L). However, the grid alone without the reference line (G) does not appear to shorten completion time.

The stereoscopic display results in figure 8 show that mean completion times with stereoscopic display are short over all visual conditions, regardless of the presence of a grid or reference lines. Especially, stereoscopic display data in columns I and II show that stereoscopic displays maintain fast performance even with extreme elevation angles. Comparable mean completion times between monoscopic and stereoscopic displays in columns III, IV, and V demonstrate that pick-and-place performance with monoscopic perspective displays, if reference lines are provided and suitable perspective parameters are chosen, can be as good as that with stereoscopic displays.

## Comparison With Three-Axis Manual Tracking Tasks

It is observed that the mean-completion-time plots obtained from the pick-and-place experiments in this paper are quite similar to the normalized rms tracking error plots obtained from the three-axis manual tracking experiments in reference 2. This strong similarity suggests that the results obtained in this paper are not task-specific, but may be applicable to other tasks.

## Choice of Display

There are many kinds of depth cues that a display can provide. Monoscopic display can provide monocular depth cues such as interposition (occlusion), brightness (light and shade), perspective projection (size), and monocular motion parallax. The human operator's knowledge and learning can also provide strong depth information pertaining to a 3-D model of a working environment. Stereoscopic display also provides a distinct binocular depth cue, called stereo disparity or binocular parallax. Consideration of these cues basically explains the experimental results of Pepper, Smith, and Cole.[18] Their results indicated that stereoscopic display performance was superior to monoscopic display performance under most conditions tested, although the amount of improvement varied with task, visibility, and learning factors. For some simple telemanipulation tasks, monocular depth cues and cognitive depth cues from knowledge and learning may be enough for successful and reliable performance, and there will be no advantage in using stereoscopic display.[15] However, for some complex tasks, monocular and cognitive depth cues may be insufficient or unavailable for successful performance with monoscopic display, and the use of stereoscopic display could significantly enhance performance. In our experiments, monocular depth cues were minimized, and target positions were randomly arranged to minimize learning effect. Consequently, our experimental results showed that pick-and-place performance with stereoscopic display was superior to monoscopic display when visual-enhancement depth cues were not presented.

Our results also showed that when reference lines were presented for visual enhancement, monoscopic display performance with adequate perspective parameters was equivalent to stereoscopic display performance. In order to present reference lines on the monoscopic display, 3-D position information of the displayed objects must be available. In a graphic display of current manipulator and camera positions, 3-D position information is normally available via joint position sensors, and reference lines can be easily provided. In a television image display of the working environment, only camera views are normally available for 3-D position information. Under current technology, a machine vision system that extracts 3-D position information of each pixel in real time from a stereo camera view is too difficult to construct,[25] although the human visual system can easily produce a 3-D image from a stereoscopic view. However, a special-purpose, machine-vision system that extracts 3-D position information of only some salient points in real time from a stereo camera view can be built. Then, reference lines for these points can be presented or superimposed on the monoscopic television display for enhanced teleoperation.

## CONCLUSION

Results of the perspective parameter experiments indicate that in order to attain good performance with a monoscopic perspective display, adequate parameter values should be chosen. For

example, extreme elevation angles or excessive azimuth angles result in very long mean completion times. Results of the visual-enhancement experiment indicate that the horizontal grid does not appear to improve pick-and-place performance in our task. The vertical reference line, however, was significant in improving performance with monoscopic perspective display. When the monoscopic display was defined with appropriate perspective parameters and provided with adequate visual-enhancement depth cues such as reference lines, the monoscopic display allowed pick-and-place performance equivalent to that of the stereoscopic display. Stereoscopic display showed short mean completion times over all visual display conditions regardless of the presence of the grid or the reference lines.

Strong similarities were observed between the mean completion time results of the three-axis pick-and-place experiments for various display conditions and the normalized rms error results of the three-axis manual tracking experiments reported previously. This demonstrates that the effects seen are robust and not task-dependent.

## Acknowledgment

# REFERENCES

[1]E. G. Johnsen and W. R. Corliss, Teleoperators and Human Augmentation, NASA SP-5047, 1967.

[2]E. G. Johnsen and W. R. Corliss, Human Factors Applications in Teleoperator Design and Operation, Wiley-Interscience, 1971.

[3]E. Heer, Remotely Manned Systems: Exploration and Operation in Space, California Institute of Technology, 1973.

[4]J. Vertut and P. Coiffet, Teleoperations and Robotics: Evolution and Development, Robot Technology Volume 3A, Prentice-Hall, 1986.

[5]A. K. Bejczy, "Sensors, Controls, and Man-Machine Interface for Advanced Teleoperation," Science, Vol. 208, No. 4450, pp. 1327-1335, 1980.

[6]J. Vertut, R. Fournier, B. Espiau, and G. Andre, "Sensor-aided and/or Computer-aided Bilateral Teleoperator System (SCATS)," RO.MAN.SY Proc. of the 5th Symposium (also in Theory and Practice of Robots and Manipulators, ed. by A. Morecki, G. Bianchi, and K. Kedzior, pp. 281-292, 1984).

[7]W. R. Ferrell and T. B. Sheridan, "Supervisory Control of Remote Manipulation," IEEE Spectrum, Vol. 4, No. 10, pp. 81-88, 1967.

[8]M. A. Bronez, M. M. Clarke and A. Quinn, "Requirements Development for a Free-Flying Robot—The Robin," Proc. of IEEE Robotics and Automation, Vol. 1, pp. 667-672, 1986.

[9]L. M. Jenkins, "Telerobotic Work System—Space Robotics Application," Proc. of IEEE Robotics and Automation, Vol. 2, pp. 804-806, 1986.

[10]M. D. Montemerlo, "NASA's Automation and Robotics Technology Development Program," Proc. of IEEE Robotics and Automation, Vol. 2, pp. 977-986, 1986.

[11]M. M. Clarke and M. A. Bronez, "Telerobotics for the Space Station," Mech. Eng., Vol. 108, No. 2, pp. 66-72, 1986.

[12]K. Leinemann and E. G. Schlechtendahl, "Computer Graphics Support for Remote Handling Simulation and Operation," Trans. Am. Nucl. Soc., Vol. 46, pp. 771-772, 1984.

[13]A. K. Bejczy, R. S. Dotson, J. W. Brown, and J. L. Lewis, "Manual Control of Manipulator Forces and Torques using Graphic Display," IEEE Proc. of the Int. Conf. on Cybernet. and Society, pp. 691-698, 1982.

[14]A K. Bejczy, J. W. Brown, and J. L. Lewis, "Evaluation of Smart Sensor Displays for Multidimensional Precision Control of Space Shuttle Remote Manipulator," 16th Ann. Contr. Conf. on Manual Contr., pp. 607-627, 1980.

[15]W. N. Kama and R. C. DuMars, "Remote Viewing: A Comparison of Direct Viewing, 2-D and 3-D Television," AMRL-TDR-64-15, Wright-Patterson Air Force Base, 1964.

[16]P. Rupert and D. Wurzburg, "Illumination and Television Considerations in Teleoperator Systems," in Remotely Manned Systems, ed. by E. Heer, pp. 219-228, California Inst. Technol., 1973.

[17]L. A. Freedman, W. H. Crooks, and P. P. Coan, "TV Requirements for Manipulation in Space," Mechanism and Machine Theory, Vol. 12, No. 5, pp. 425-438, 1977.

[18]R. L. Pepper, D. C. Smith, and R. E Cole, "Stereo TV Improves Operator Performance Under Degraded Visibility Conditions," Opt. Eng., Vol. 20, No. 4, pp. 579-585, 1981.

[19]R. L. Pepper, R. E. Cole, E. H. Spain, and J. E. Sigurdson, "Research Issues Involved in Applying Stereoscopic Television to Remotely Operated Vehicles," Proc. SPIE, Vol. 402—Three-Dimensional Imaging, pp. 170-173, 1983.

[20]W. S. Kim, S R. Ellis, M. Tyler, and L. Stark, "Visual Enhancement for Telerobotics," IEEE 1985 Proc. Int. Conf. on Cybern. and Society, pp. 819-823, 1985.

[21]W. S. Kim, S. R. Ellis, B. Hannaford, M. Tyler, and L. Stark, "A Quantitative Evaluation of Perspective and Stereoscopic Displays in Three-Axis Manual Tracking Tasks," IEEE Trans. Syst., Man, Cybern., Vol. SMC-17, No. 1, pp. 61-72, 1987.

[22] J. D. Foley and A. Van Dam, Fundamentals of Interactive Computer Graphics, Addison-Wesley, 1983.

[23]I. Carlbom and J. Paciorek, "Planar Geometric Projections and Viewing Transformation," Computing Surv., Vol. 1, No. 4, pp. 465-502, 1978.

[24]E. F. Pischel and J. J. Pearson, "Image Processing and Display in Three Dimensions," Proc. SPIE, Vol. 528—Digital Image Processing, pp. 23-28, 1985.

[25]M. Drumheller and T. Poggio, "On Parallel Stereo," Proc. of IEEE Int. Conf. on Robotics and Automation, Vol. 3, pp. 1439-1448, 1986.

Figure 1.— The experimental setup.

left
retinal
image

right
retinal
image

Z   Y
←   ⊙

X ↓

P

R

Q

S

retina     pupil     binocular     picture          image
           &         converging    plane            plane
      crystalline    lens          (display
           lens                    screen )

Figure 2.– Stereoscopic display.
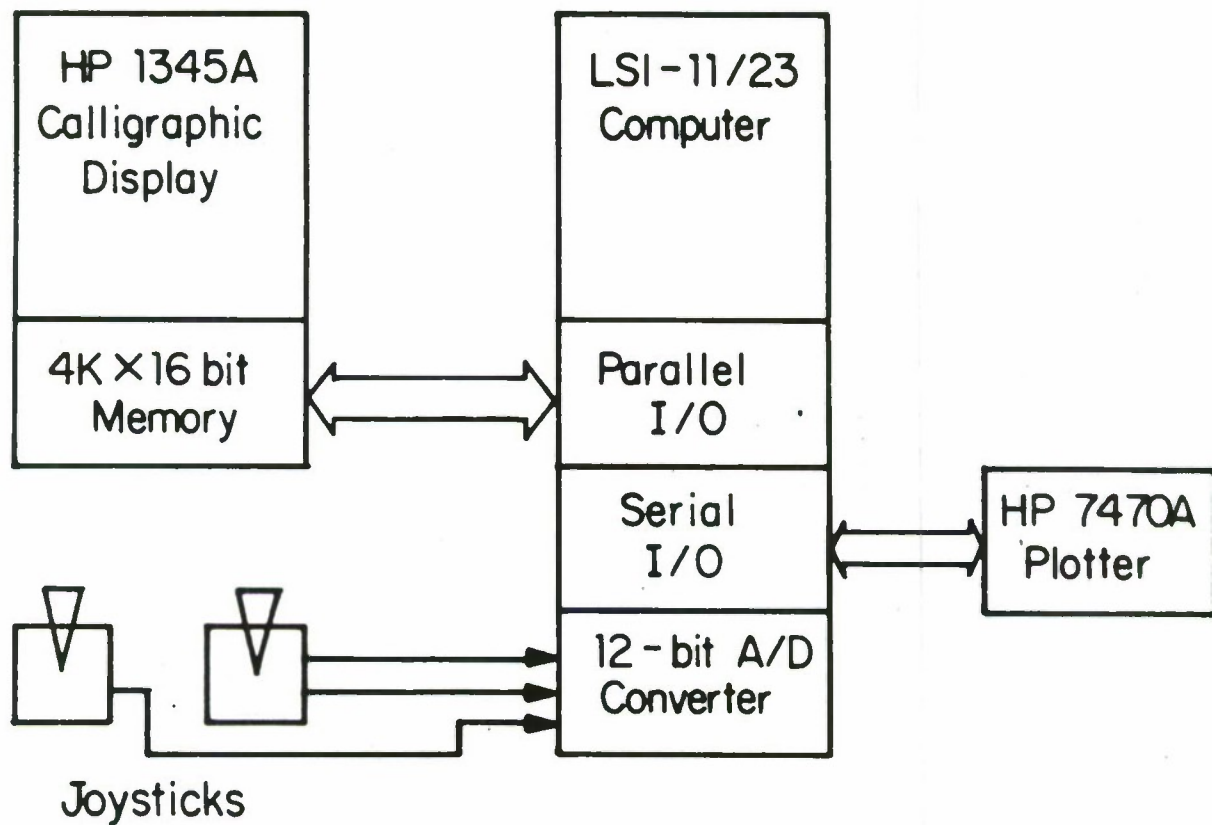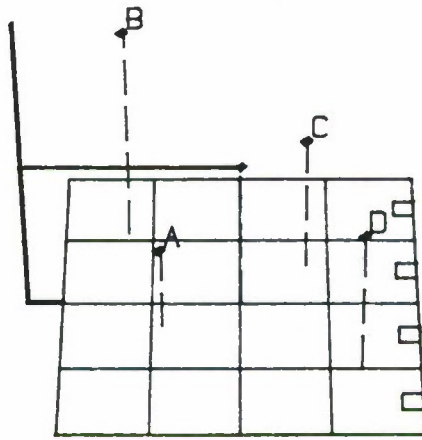
Figure 3.– A monoscopic perspective presentation using nominal perspective parameters.
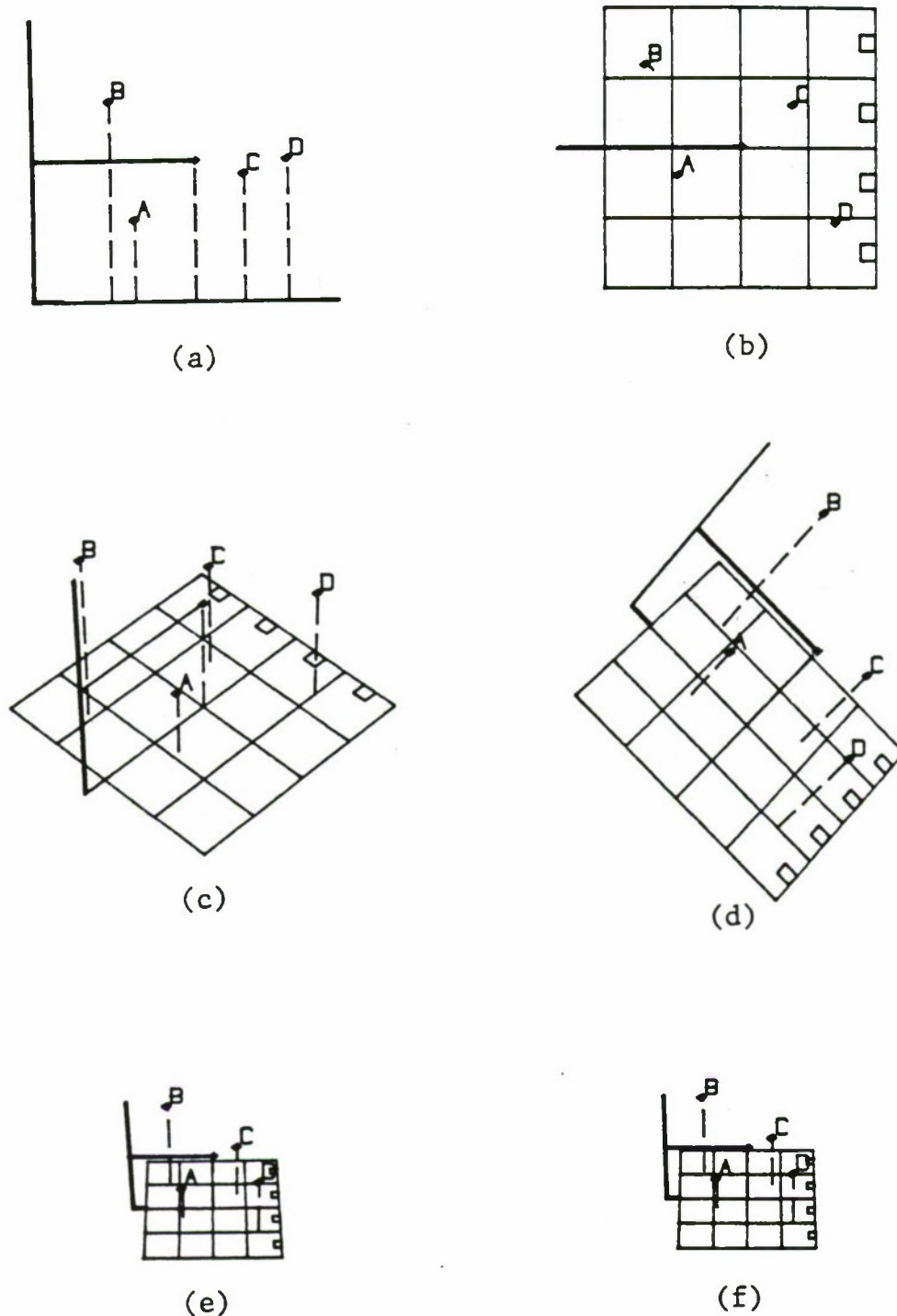
Figure 4.– Examples of various monoscopic perspective presentations with (a) an extreme 0° elevation angle, (b) the other extreme –90° elevation angle, (c) 45° azimuth angle, (d) 45° roll angle, (e) fov angle doubled to 24°, and (f) object distance doubled to 80 cm. A 5-line-by-5-line horizontal grid and vertical reference lines are presented.
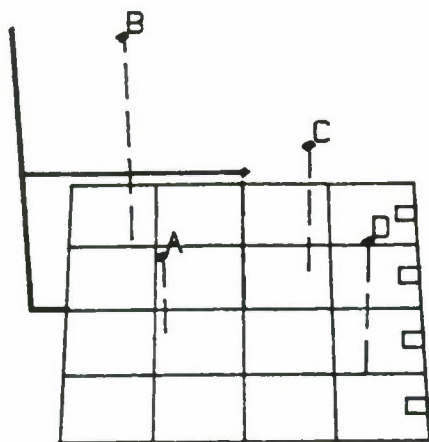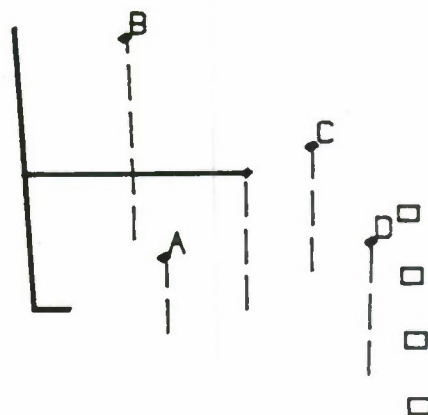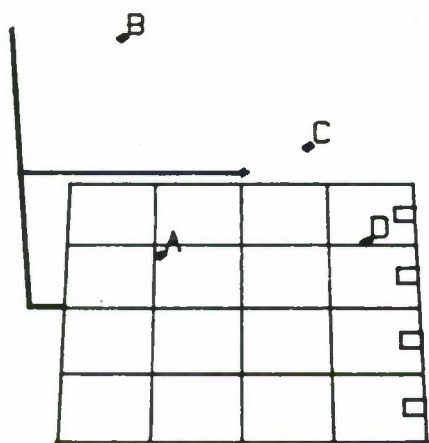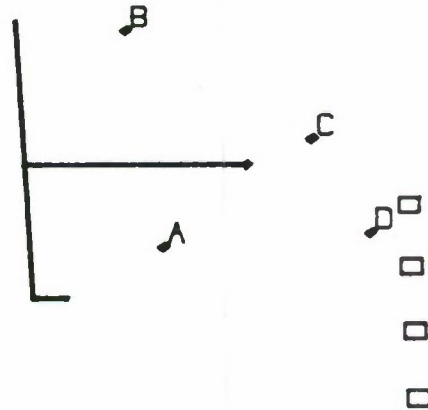
Figure 5.– Monoscopic presentations under four visual-enhancement conditions: (a) GL (presence of both grid and reference line), (b) L (reference line only), (c) G (grid only), and (d) O (neither).
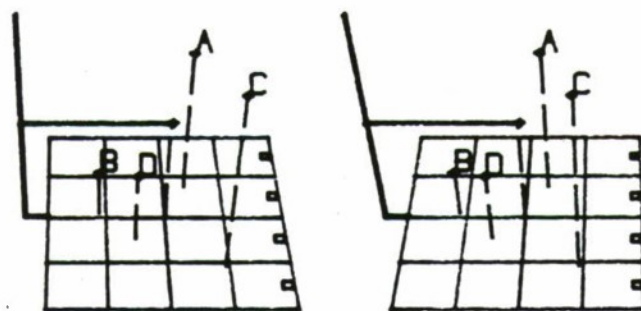
Figure 6.– An example of a stereoscopic presentation.
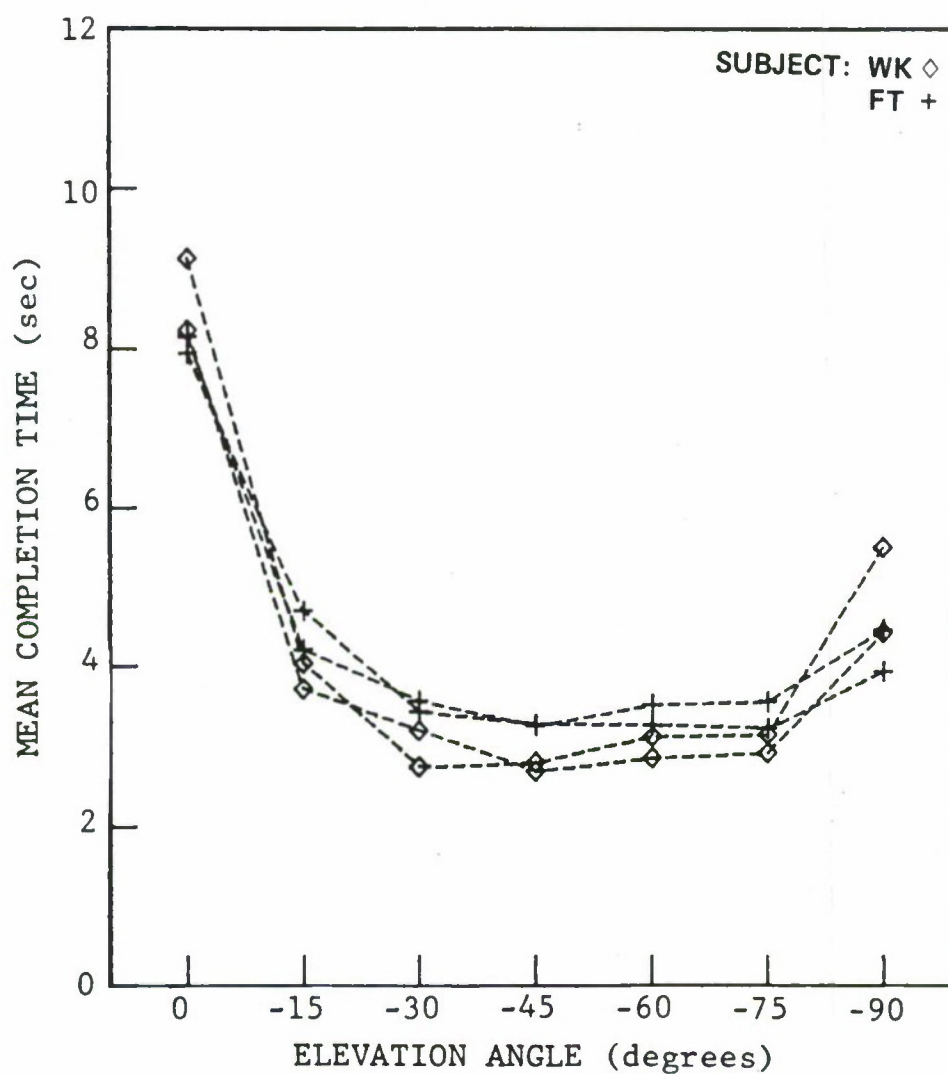
(a) Mean completion time as a function of elevation.

Figure 7.– Perspective parameter experiment. Three-axis pick-and-place performance with various monoscopic perspective displays.
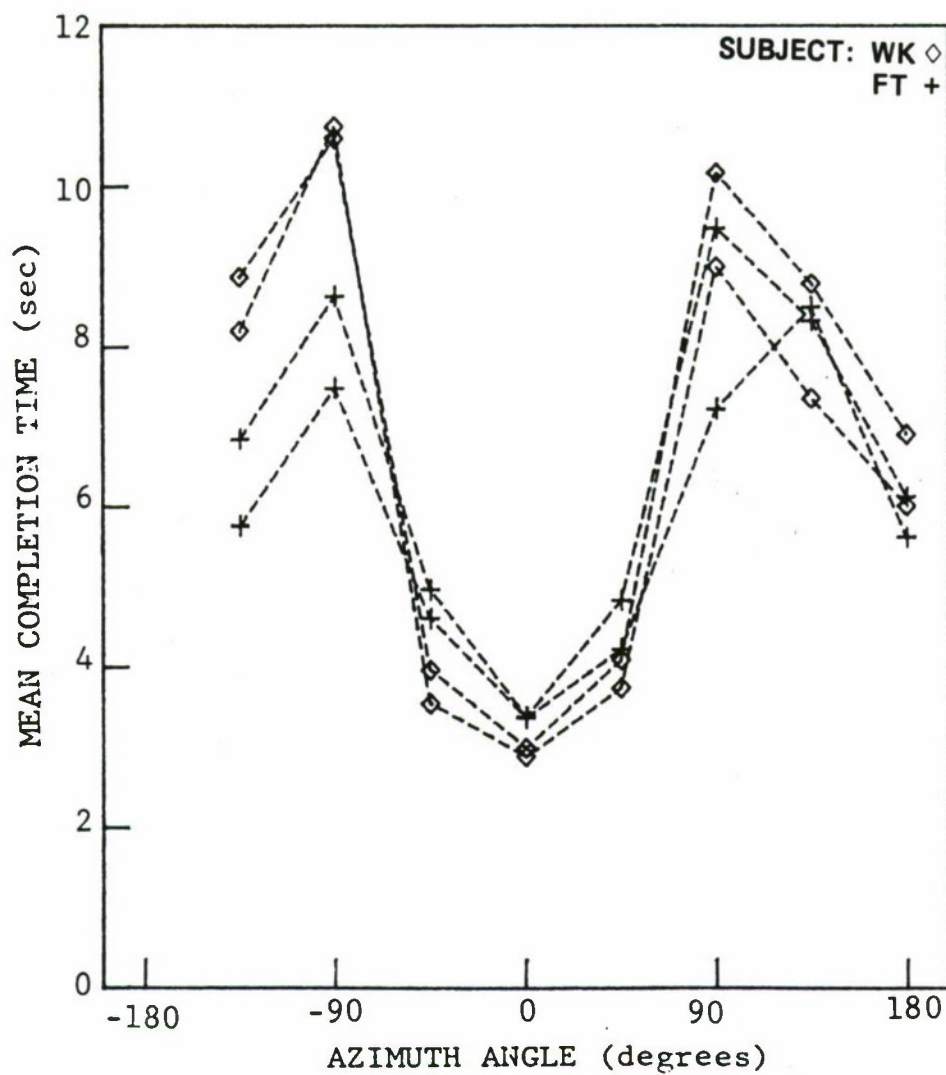
(b) Mean completion time as a function of azimuth.

Figure 7.– Continued.
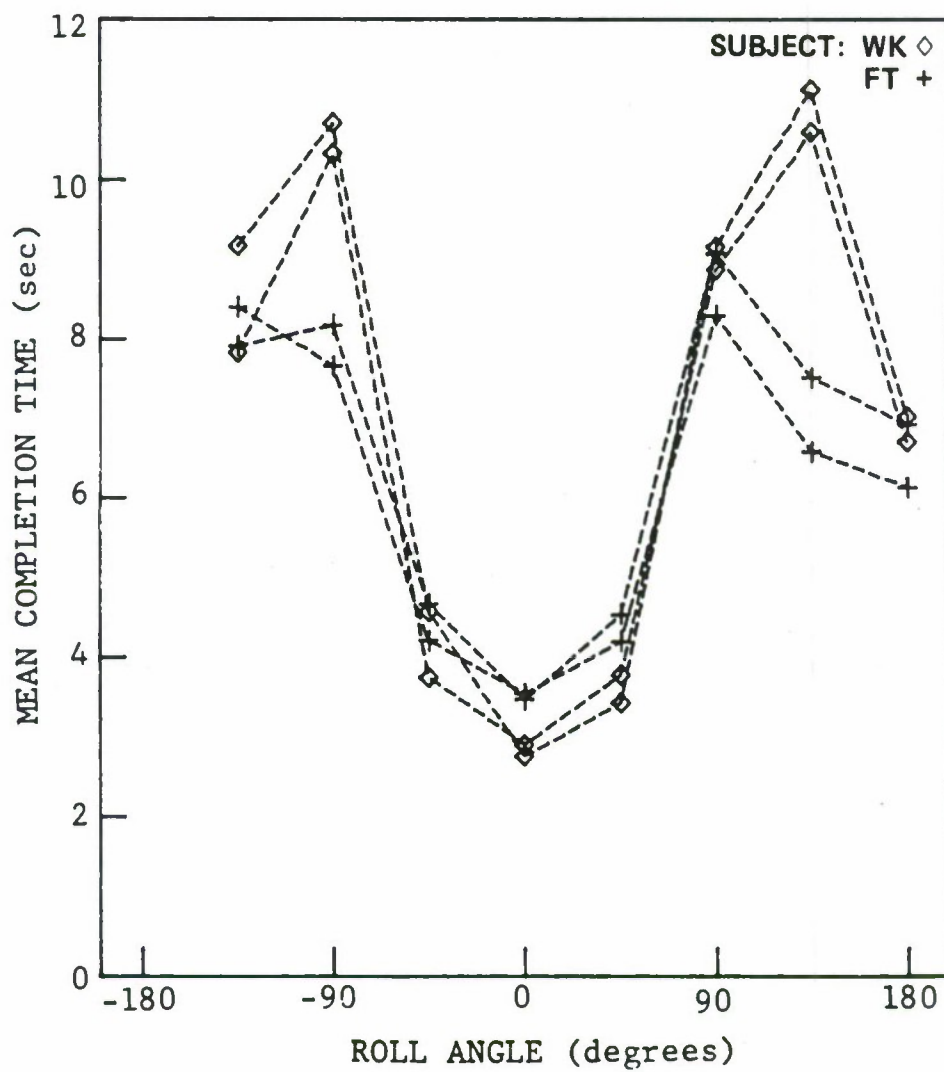
(c) Mean completion time as a function of roll.

Figure 7.– Continued.
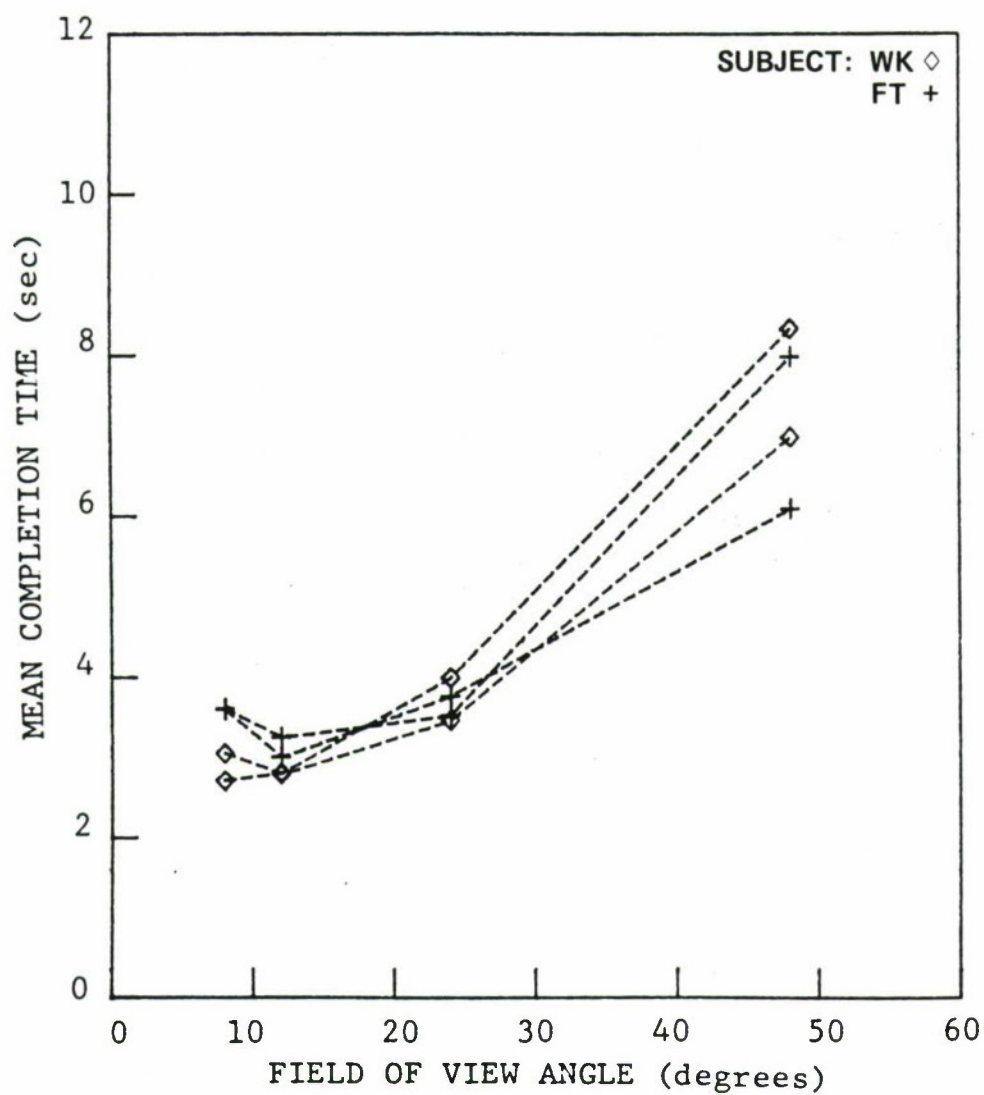
(d) Mean completion time as a function of fov angle.

Figure 7.– Continued.
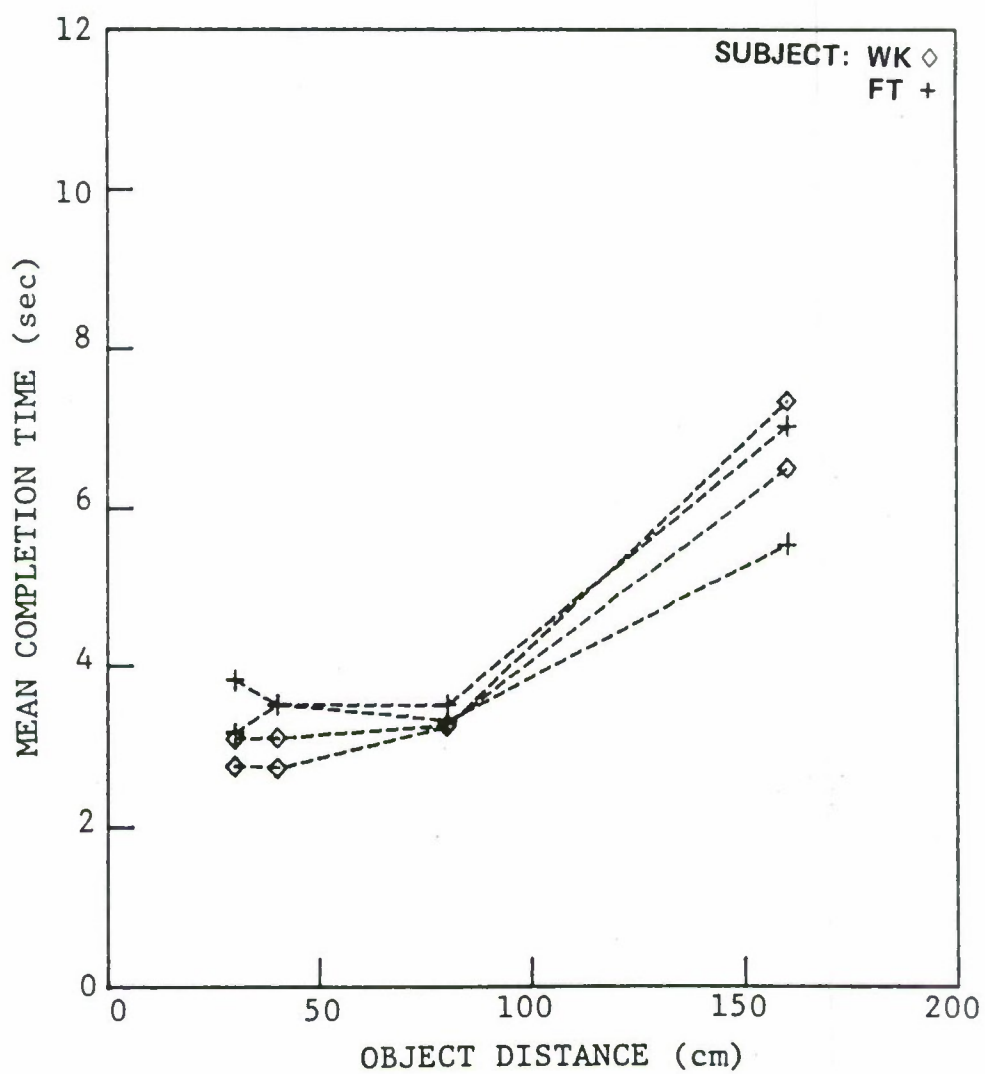
(e) Mean completion time as a function of object distance.

Figure 7.– Concluded.
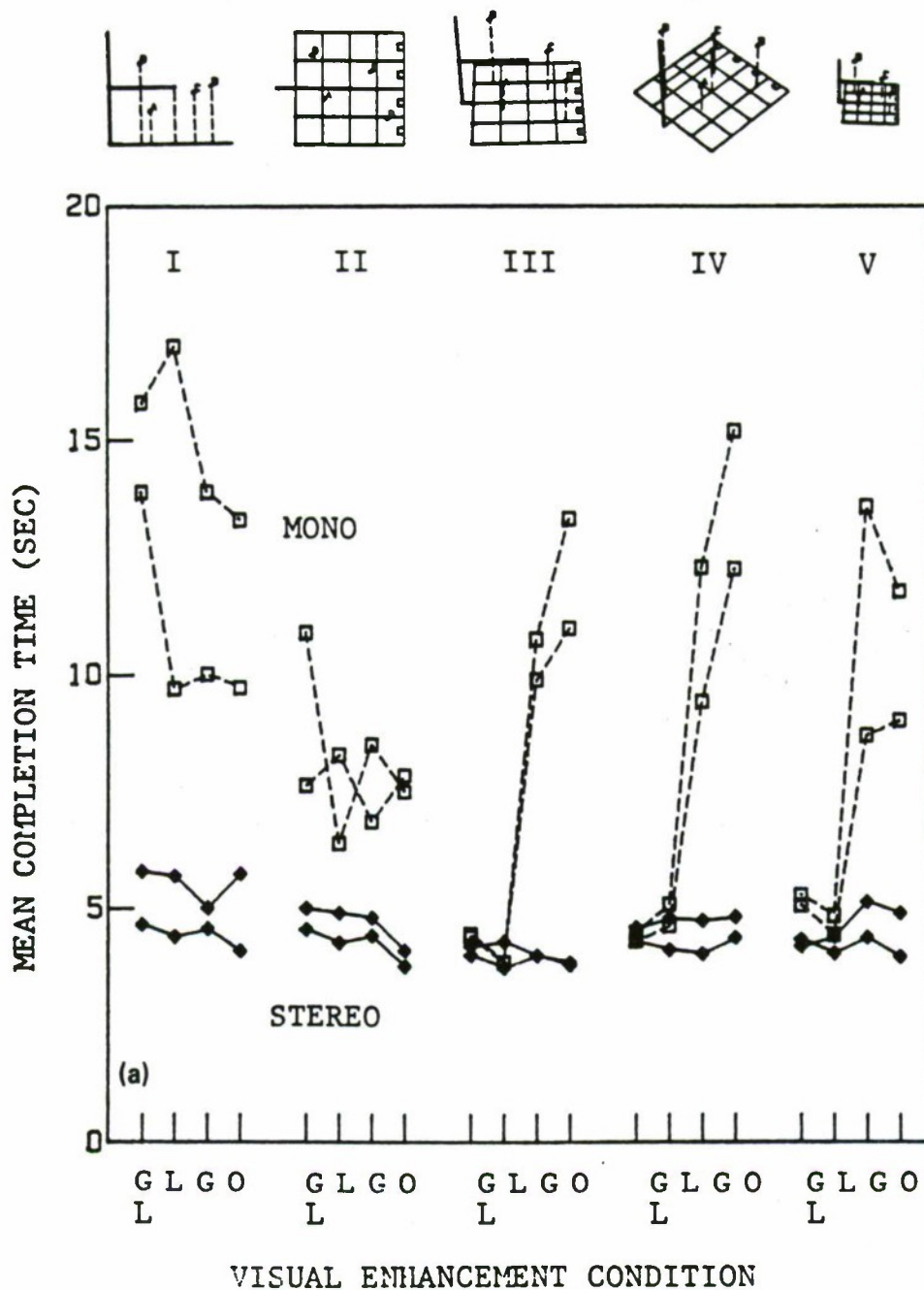
Figure 8.– Visual-enhancement experiment. Three-axis pick-and-place performance for four visual-enhancement conditions at each of five different perspective parameter conditions with both monoscopic display and stereoscopic display. The monoscopic presentations for the five perspective parameter conditions are shown above the plot (a). Four visual-enhancement conditions are GL (presence of both grid and reference line), L (reference line only), G (grid only), and O (neither). Subjects: (a) WK, (b) MT. Two runs for each subject. In plot (b), confidence intervals at the 95% level are shown about the means.
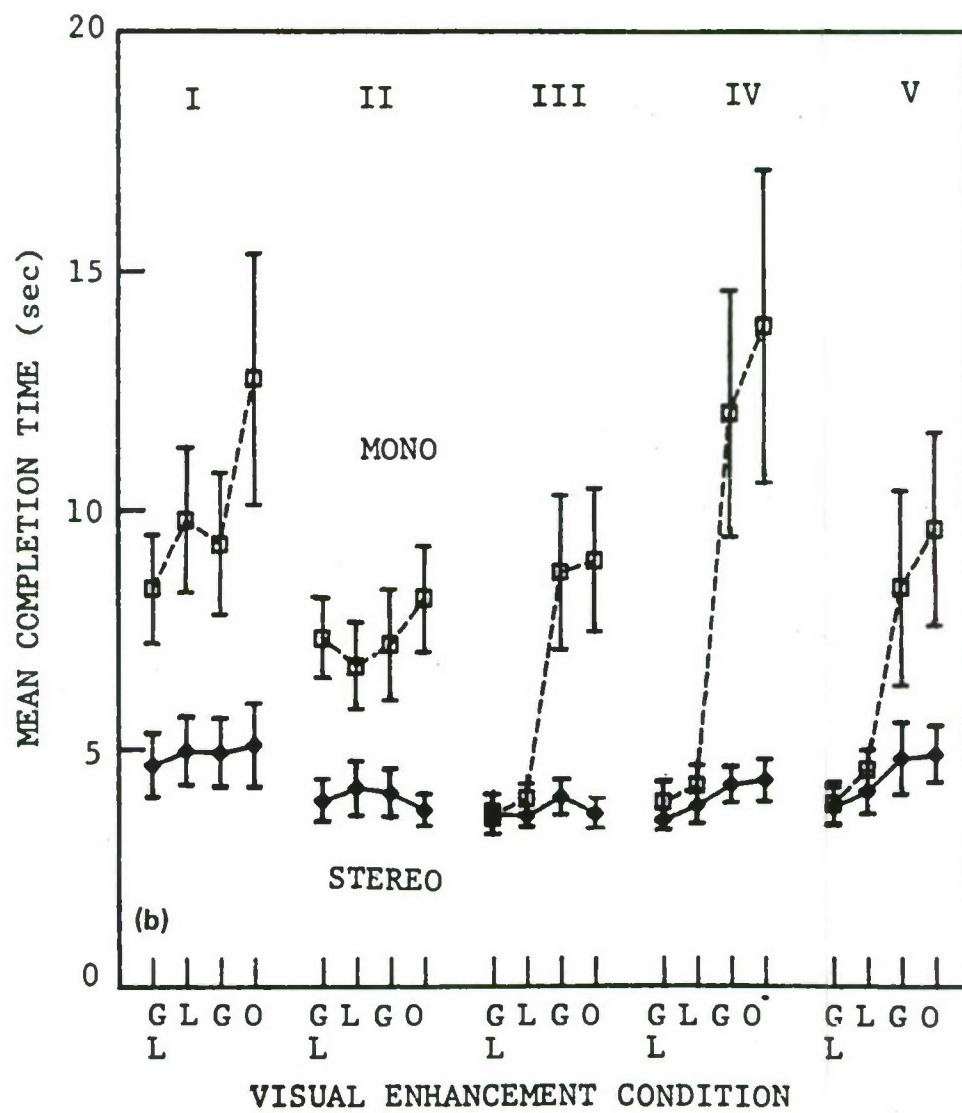
Figure 8.— Concluded.

# DIRECTION OF MOVEMENT EFFECTS UNDER TRANSFORMED VISUAL/MOTOR MAPPINGS

H. A. Cunningham and M. Pavel
Stanford University
Stanford, California

## SUMMARY

Performance in a discrete aiming task was compared under several transformed visual/motor mappings: rotations by 45°, 90°, 135°, and 180° and reflections about the horizontal and the vertical midlines. Eight aiming targets were used, corresponding to eight directions of movement: up, down, right, left, up-right, down-left, up-left, and down-right. Direction of movement effects were characterized in terms of separable visual and motor direction components, and two kinds of direction of movement effects were considered. First, a direction of movement effect paralleling that seen in rapid aiming under the usual nontransformed mapping might be seen. If it is seen for motor directions, but not visual directions, then this supports a motor factor hypothesis for the effects seen under the nontransformed mapping. Second, because rotations, but not reflections, are physically realizable two-dimensional (2-D) transformations, a visual/motor control system which is sensitive to physical constraints should perform reflections, but not rotations, in a piecemeal fashion. Results supported the hypothesis that a motor factor having to do with complexity of limb movement accounts for differences in movement accuracy between right and left oblique directions. Direction of movement effects were more evident in reflections than in rotations, and were consistent with the hypothesis that the visual/motor-control system seeks a physically realizable 2-D rotation solution to reflections. Results also suggested that reversal of two orthogonal basis dimensions is far less difficult than reversing only one and leaving the other intact.

## INTRODUCTION

This research investigates directional nonuniformities in the performance of a 2-D discrete aiming task, under transformed mappings between visual and motor spaces. Various rearrangements of the visual/motor map have been studied over the years (see Howard, 1982, for an excellent review). This work has focused primarily on the process of adaptation to visual/motor transformations. The present research, in contrast, compares the effects of different transformations and examines direction of movement effects within and between different transformations.

Direction of movement effects (DMEs) have important implications for our understanding of human visual/motor control. If there is nonuniformity in performance under physically uniform conditions, this reveals something about the organization of the internal representation of external space and about the mechanisms involved in visual/motor control. DMEs also are of practical importance because they can lead to biases in an operator's input to a system. Such biases are not easily detected when evaluating overall performance of the task, because they involve only a subset of the inputs. Understanding this source of bias would allow the development of systems that prevent biases or correct for them during operation.

In this research, visually guided aiming has been studied under two kinds of transformation of the usual directional mapping between a horizontal input surface (motor space) and a vertical display screen (visual space), which is such that:

$$Right \rightarrow Right$$
$$Left \rightarrow Left$$
$$Forward \rightarrow Up$$
$$Backward \rightarrow Down$$

This mapping is a natural one that humans as young as 3 yr of age can do immediately without any period of adaptation. This mapping will be referred to as the "usual" or "nontransformed" mapping.

The transformations that were studied constitute a subset of the linear orthogonal transformations. They were 1) rotations about the center of the space and 2) reflections about axes in the space. Rotations and reflections both preserve line length, angles, and parallelism of points in the original space when mapped into corresponding points in the image space. In general, the expression:

$$TX = X'$$

describes a transformation, T, of points $X = [x \ y]$ in the original space into points $X' = [x' \ y']$ in the image space. In this research, T took one of the following forms:

$$T_{ROT} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

$$T_{HREF} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$T_{VREF} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

$$T_{OBREF} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

These transformations represent, respectively, rotation about the center of the 2-D space by angle q, reflection about the horizontal midline of the space, reflection about the vertical midline, and reflection about a 45° line going through the center of the space.

## METHODS

Six right-handed subjects performed a discrete aiming task with multiple possible target positions. The visual display was a vertical CRT screen and the motor input was movement of a hand-held stylus on a horizontal digitizing tablet. There were eight possible target positions, arranged at 45° intervals around the center. An aiming trial consisted of 1) the subject aligning the cursor with a marker at the center of the display screen; 2) a cueing tone sounding; 3) after a variable foreperiod (250 to 750 msec), a target appearing in one of the target positions; 4) the subject capturing the target by moving the cursor into alignment with it; and 5) the target extinguishing. Subjects were instructed to emphasize accuracy over speed and to execute as straight a trajectory as possible on every trial.

Each experimental session consisted of 32 baseline trials under the usual mapping, followed by 128 trials under one of the six transformed mappings: rotation of 45°, 90°, 135°, or 180°, or reflection about the vertical midline or about the horizontal midline. Transformations of the motor space relative to the visual space were effected using a combination of software manipulation and physical rotation of the digitizing tablet.

Root-mean-squared error (RMS ERROR) measured the deviation of a trajectory from a straight line and is reported here as the measure of difficulty experienced by subjects under the various visual/motor mappings. Reaction time and angular error of the initial segment of the trajectory were also obtained on each aiming trial, and are reported elsewhere (Cunningham, 1987a).

## HYPOTHESES

Two aspects of DMEs were considered, and they correspond to two specific questions that were asked. First, can DMEs observed under transformed mappings help us to understand DMEs observed under nontransformed mappings? Previous work by this author (Cunningham, 1987b) has shown that under the nontransformed mapping, movement in some directions produces more error than movement in others. Specifically, among right-handed subjects movement along the left oblique produces more error than movement along the right oblique, and horizontal movement produces more error than vertical movement. Are these directional nonuniformities due to properties of the motor system or to nonmotor properties of visual or cognitive processes? Under the nontransformed visual/motor mapping, motor direction and nonmotor direction are congruent (i.e., confounded). Testing left-handed subjects will not disconfound them because it is possible that left-handers have reversed lateralization of information processing at many levels, not just in the motor system. Transformation of the visual/motor mapping, however, allows us to disconfound motor and nonmotor factors because directions of movement are no longer aligned in the usual way. Under a 90° rotation, for example, the visual right oblique becomes the motor left oblique, and vice versa. Under a 135° rotation, visual vertical corresponds to motor right oblique, and so forth. Thus it was asked: Will the expected pattern of DMEs be observed under transformed visual/motor mappings and, if so, will it be observed in display directions only (visual), in tablet directions only (motor), or in both?

The second question arises from considerations of the properties of the two kinds of transformations studied in this research: rotations and reflections. These are both linear orthogonal

transformations, and so are mathematically similar. They differ, however, in one important respect: they are not equally physically realizable operations. A rotation of points on a 2-D surface is a rigid motion that can be realized in two dimensions. A reflection of points on a 2-D surface, however, is neither rigid nor physically realizable in two dimensions. Are the mechanisms responsible for visual/motor control sensitive to this difference? If so, performance under reflections should be qualitatively different from that under rotations. Specifically, it was asked: Is directional nonuniformity more likely to occur under reflection than under rotation as the system seeks a physically realizable solution to the transformation?

## RESULTS

Transformation condition exerted an important influence on aiming error. On average, the four rotations differed both from one another and from the reflections. The condition which produced the highest average RMS ERROR was the 90° rotation. This was followed by the two reflections (which were the same) and the 135° rotation. The 45° and 180° rotations produced the least error and were similar to one another. These averages are for all movement directions under a particular transformation, and they are consistent with results obtained by other investigators in a three-dimensional tracking task under transformed visual/motor mappings (Kim et al., 1987). DMEs were also seen under both kinds of transformation, but they were qualitatively different under rotation and reflection. In figure 1, RMS ERROR under the four rotation conditions is plotted against axis of movement: horizontal (right and left), vertical (up and down), right oblique (up-right and down-left), and left oblique (up-left and down-right). Axes of movement correspond to directions of movement on the tablet (motor direction), irrespective of display direction. The vertical offset of the curves for each condition indicates the overall effect of the transformation condition. The expected right oblique/left oblique difference is seen for the 90° and 135° rotations. This is also true for the 45° condition, although the scale of this plot makes the difference less obvious. The horizontal-vertical difference seen under nontransformed mapping was *not* preserved in either visual or·motor coordinate systems under rotation.

An interesting and very different pattern of DMEs emerges under the reflection conditions. Figure 2 shows RMS ERROR under a reflection about the horizontal midline. Note that under this transformation, the horizontal axis (axis of reflection) is preserved: direction of travel along the axis is the same as under the nontransformed mapping. The vertical axis is reversed. The right and left obliques are exchanged, which is equivalent to rotating each of them by 90°. The surprising result shown in this figure is that the axis along which sign is preserved (right and left) has considerably higher aiming error than that along which sign is reversed (up and down). The axes corresponding to 90° rotations also exhibit high error.

Figure 3 demonstrates that this effect is also seen under the reflection about the vertical axis. Here, vertical axis movement is preserved as in the nontransformed mapping, and the error for movements along that axis is high. The horizontal axis is reversed, and error for movements along that axis is low. Again, error for movements along the other two axes is also high. The significance of direction of movement under reflections appears to relate not to the orientation of a movement axis in external space, but rather to its orientation with respect to the transformation performed on the space.

# PRELIMINARY DISCUSSION

DMEs were observed under both rotation and reflection transformations. Under rotation, the pattern of results for right versus left oblique confirmed a probable motor locus for the right oblique advantage. This was seen in three out of the four rotation transformations and was especially strong in those where the overall error is high (90° and 135° rotations). This motor effect is consistent with the fact that movement along the right oblique can be done with arm movements from the elbow, whereas movement along the left oblique requires movement from the shoulder. Movement from the shoulder involves more joints and the control of more mass than does movement from the elbow.

The DMEs seen under reflection are qualitatively different from those seen under rotation. They are also large. Under reflection, the reversed axis has the lowest aiming error, and the two oblique axes have the highest. The error along the axis of reflection was surprisingly high, considering that the reflection transformation preserves that axis entirely. To what may we attribute these directional nonuniformities seen under reflection? There are two separate questions to answer:

1. Why do the oblique axes exhibit higher error than the nonoblique axes? Is it because they are oblique or because they are transformed by the equivalent of a 90° rotation?

2. Why do the preserved axes exhibit greater error than the reversed axes?

## Another Transformation:   Oblique Reflection

In order to answer the first question, an additional condition was run: reflection about an oblique axis. Under this reflection, the right oblique was the axis of reflection and so was preserved. The left oblique was thus reversed. The horizontal and vertical axes were exchanged for one another, which is equivalent to a 90° rotation of each of them. Figure 4 shows the result of this reflection. Observed DMEs are consistent with those found under horizontal- and vertical-axis reflection. The reversed axis exhibits low error and the preserved axis exhibits high error. The axes whose transformation is equivalent to a 90° rotation also exhibit high error.

## GENERAL DISCUSSION

DMEs were observed under several different transformations of the usual mapping between visual (display) space and motor (input) space. Two types of DMEs were seen. First, aiming error was lower for right oblique motor directions than for left oblique motor directions, irrespective of visual direction. This supports the hypothesis that the right oblique "advantage" seen under nontransformed visual/motor mapping is due to motor factors. A tendency for vertical error to be lower than horizontal error under the nontransformed mapping was *not* seen in either the motor or the visual directions under the transformed mappings.

DMEs also differed qualitatively between rotations, on the one hand, and reflections, on the other. Under reflection, DMEs are related to an axis of movement's orientation with respect to the

axis of reflection, not with respect to external space. The fact that human performance exhibits this particular kind of directional nonuniformity under reflection, but not under rotation, is consistent with the hypothesis that the human representation of 2-D space is constrained by physical realizability. The pattern of DMEs under reflection suggests that the human imposes a 2-D rotation solution on the reflection condition. Axes whose transformation is equivalent to a 180° rotation exhibit less error than those whose transformation is equivalent to a 90° rotation, just as a 180° rotation of the entire space produces less error, in all directions, than a 90° rotation of the entire space.

Another interesting aspect of the DMEs found under reflection (and one which complicates somewhat the 2-D solution hypothesis) is the strong tendency for the reversed axes to exhibit lower error than the nonreversed axes. This was seen in every reflection. This is probably due to error correction during movement execution. During execution of a movement, subtle corrections are required to keep the trajectory on a straight path toward the target. For a straight-line trajectory, corrective movements will have a large vector component in the dimension orthogonal to the straight-line path. Under reflection, when moving along the axis of reflection (the preserved axis), the orthogonal dimension is reversed. The small, quick, and largely automatic corrections made during movement execution will initially be in the wrong direction. As the error is detected, further automatic attempts to correct it may result in enhancing it instead. This is equivalent to reversing the sign of a feedback loop and the result is similar: error "blows up." In the case of movement along the reversed dimension, the orthogonal dimension (dimension of correction) is preserved and so automatic corrections reduce the error as they should.

In summary, DMEs are intrinsic to discrete aiming on a 2-D surface. The mechanisms responsible for visual/motor control are sensitive to motor factors having to do with the number of joints involved in movement in a given direction. They also appear to be constrained to find 2-D physically realizable solutions to visual/motor transformations, even when these solutions do not exist.

## REFERENCES

Cunningham, H. A. (1987a). Development of visual-motor control. Submitted for publication to J. Exp. Psychol: Human Percept. Perf.

Cunningham, H. A. (1987b). Human motor control under transformed visual-motor mappings. Submitted for publication to Biol. Cybernet.

Howard, Ian P. (1982) Human Visual Orientation. Toronto: John Wiley & Sons.

Kim, W. S., Ellis, S. R., Tyler, M. E., Hannaford, B., and Stark, L. W. (1987) Quantitative evaluation of perspective and stereoscopic displays in three-axis manual tracking tasks. IEEE Trans. Syst., Man, and Cybernetics, 17, 61-72.
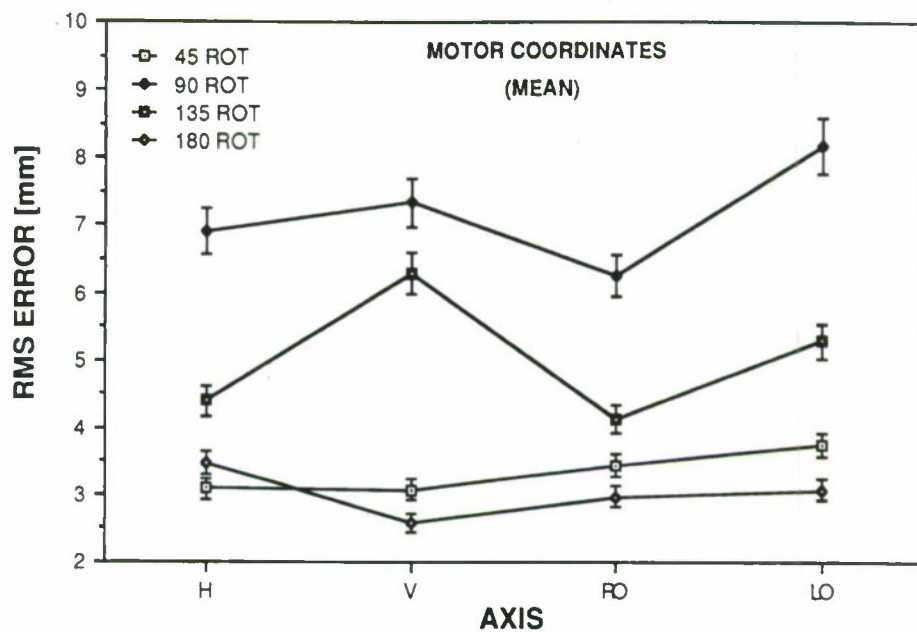
Figure 1.– RMS ERROR plotted against axis of movement in motor coordinates (directions of movement on the tablet, irrespective of display direction). Axes are horizontal, vertical, right oblique, and left oblique. Note that the right oblique/left oblique difference seen under the usual mapping is preserved in motor coordinates and so is probably motor in origin. The horizontal/vertical difference observed under the usual mapping is *not* preserved.
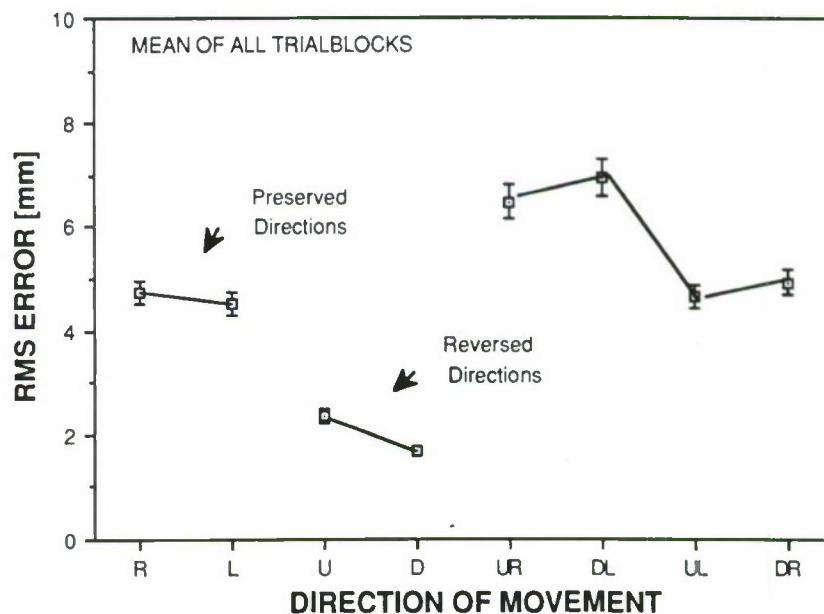


Figure 2.– RMS ERROR plotted against direction of movement for eight directions. The horizontal axis (right and left) is preserved and the vertical axis (up and down) is reversed. Oblique axes correspond to a 90° rotation. Note that the oblique axes have the highest error, the reversed axis the least, the preserved axis intermediate. Note also that the "motor oblique effect" is present (right and left obliques are exchanged).
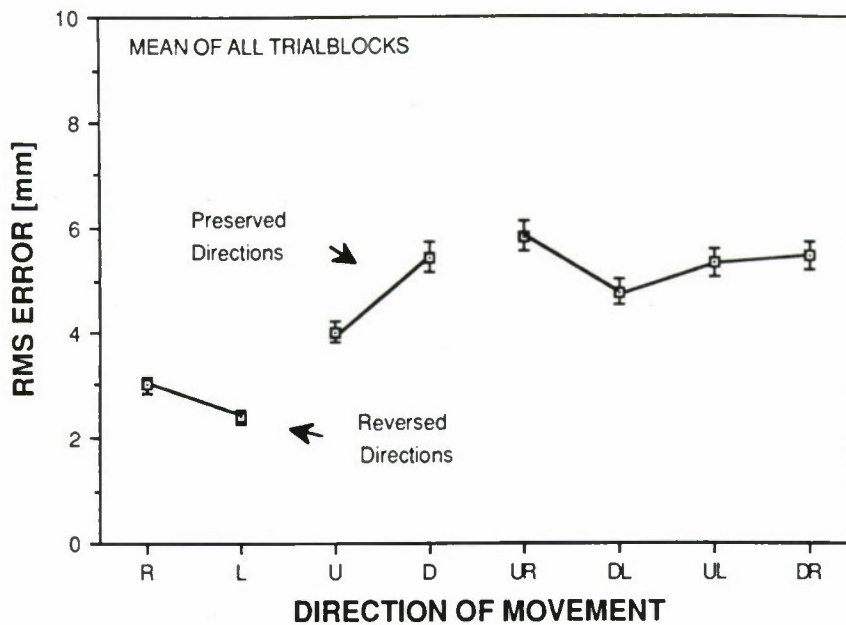
Figure 3.– RMS ERROR plotted against direction of movement under reflection about the vertical axis. The pattern of errors with respect to the oblique axes, the preserved axis, and the reversed axis is essentially the same as that seen under horizontal reflection.
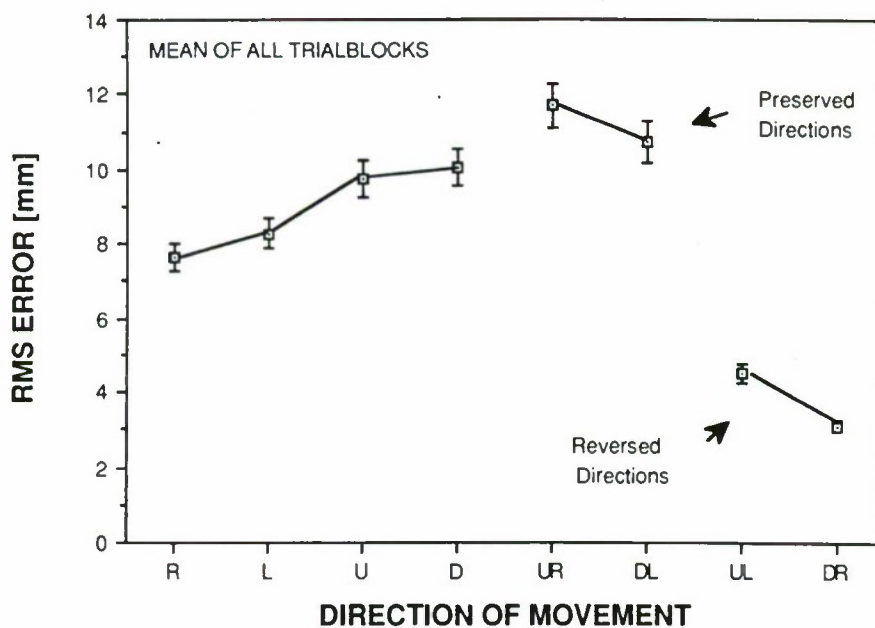


Figure 4.– Under reflection about the right oblique axis, the reversed and preserved axes are the obliques. Yet the same pattern of error is seen: reversed axis exhibits low error, and preserved axis and 90° rotation axes exhibit high error.

# DISPLAYS FOR TELEMANIPULATION

Blake Hannaford
Marcos Salganicoff
Antal Bejczy

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

## SUMMARY

Visual displays drive the human operator's highest bandwidth sensory input channel. Thus, no telemanipulation system is adequate which does not make extensive use of visual displays. Although an important use of visual displays is the presentation of a televised image of the work scene, this paper will concentrate on visual displays for presentation of nonvisual information (forces and torques) for simulation and planning, and for management and control of the large numbers of subsystems which make up a modern telemanipulation system.

## INTRODUCTION

Teleoperation consists of the control of a remote manipulator in order to perform mechanical actions usually associated with the function of the human arm and hand. This extension of manual dexterity to hostile environments requires high sensory feedback bandwidth to replicate perceptual inputs normally available to the human.

Augmented by computers and advances in robot sensor development, the application of teleoperation has been extended to the areas of deep sea, underground, and space exploration. Future space missions will require a more advanced teleoperator with automation capability to perform many new tasks including satellite retrieval or repair, space station construction, and payload handling (ref. 1).

Visual displays drive the human operator's highest-capacity input channel, allowing an important means of closing the dextrous manipulation loop. The televised image of the work scene affords the operator an important means of receiving qualitative and nonsymbolic quantitative information about the work environment. This type of display has the advantage of providing information in a natural, unencoded form, but can suffer from perspective ambiguities if any parameters such as the viewing angle, lighting conditions, display resolution, refresh rate, or reference frame are ill chosen (refs. 2 and 3). Additionally, televised display can rapidly exhaust the available transmission data rates in the downlink. Displays which represent the state variables in encoded form offer a much more efficient use of the downlink if their chosen form can be quickly decoded and easily understood by the human operator.

There are many important parameters to be displayed in telerobotic displays. Displays can provide information about the proximity of the end effector to goals and obstacles (ref. 4); the

forces and moments exerted at the wrist frame of the manipulator (ref. 5); the current configuration and work envelope of the manipulator relative to objects in the task space, including regions near manipulator singularities that should be avoided; and mass distribution of objects in the environment.

This research was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

## Force Torque Displays

A long-term effort in our laboratory has focused on the display of forces and torques arising from a remote manipulator's interaction with the environment. Visual displays complement the ability of force-feedback master manipulators when time delay, or control-station constraints preclude such aids. We have developed and evaluated several graphical formats through which this nonvisual task space information can be presented including horizontal bar graphs (ref. 5), so-called "star diagrams" (ref. 6), and various enhancements such as color coding, event-driven flags, and true perspective presentation. The star diagram display has recently been tested in over 21 hr of experimental teleoperation with resulting guidelines for future system design.

## Simulation

An Iris graphics workstation has served as a graphics engine for a number of simulation displays used for kinematic analysis of proposed telerobotic task scenarios. Examples include an animated simulation of a dual-arm, satellite-servicing task and a detailed simulation used for analysis of arm-base location and position in a dual-arm teleoperation laboratory. These perspective displays can be interactively rotated and zoomed in and out to give three-dimensional information to the operators without the problems of stereo displays. Visual enhancements such as color coding, reference grids, and manipulator work volume projections are used in place of binocular cues.

## Executive Control Displays

In a full telerobotic system, a very large number of subsystems and capabilities need to be controlled. In full systems, these will include two arms, hand controllers, and smart hands; trading and sharing of control between autonomous and telerobotic modes; and control of cameras, light sources, and other sensory systems. The traditional solution of large racks of subsystem control panels attended by a dedicated operator can be improved upon with an executive workstation which can communicate with all subsystems over a local network such as Ethernet. Recent exploration work has developed a prototype display architecture based on the desktop metaphor built into workstations such as the Macintosh or Sun. Icons representing each of the subsystems to be controlled populate a workstation screen representing the control domain. The key feature is that an operator can selectively attend to one of a large number of subsystems by selecting (clicking) its icon. The icon expands into a software control panel which displays the subsystems' status and accepts commands. Alarm conditions can be indicated on the icon to alert attention to the particular subsystem.

Telemanipulation displays are not limited to on-line situations during task execution. They can provide predictive information about the outcome of given operator-control actions if the manipulator and environmental characteristics are modeled (refs. 7 and 8), as well as a play-back for postmortem analysis of operator performance. Off-line use of all previous modes with the addition of environmental modeling allows for training of operators in routine activities with a minimum investment in hardware and low risk of damage to training facilities. The success of this approach can be seen in the widespread acceptance of flight simulators as a training tool for commercial and military pilots. Currently, one validated, high-fidelity, real-time simulator for space telemanipulation exists, the Shuttle remote manipulator simulator at the Johnson Space Flight Center (ref. 9).

This paper reports the results of display research and development at the Jet Propulsion Laboratory Advanced Teleoperator Development Laboratory in three sections: displays of force and torque data; perspective projection displays of simulated manipulators and task environments; and executive control of complex telemanipulation systems by direct manipulation.

## Force/Torque Information

When a robot manipulator interacts with the environment, forces and torques are exerted at the contact points. Information from the load cells in the robot "wrist" can be resolved into three forces and three torques representing the interaction between manipulator and environment as "felt" in the wrist. The specific coordinate system for this resolved information is arbitrary, but a useful one is to resolve the three components of force along the x, y, and z axes, and the three components of torque to the pitch (x), yaw (y), and roll (z) axes of the manipulator hand.

Although considerable attention has been focused on using backdrivable master manipulators to provide contact force information to the operator, there are cases where this direct information is insufficient or impractical. In particular, visual displays complement the ability of force-feedback master manipulators when time delay, numerical accuracy, or control-station constraints preclude such aids. Graphical displays can also indicate task-specific constraints which must be satisfied during manipulation and whether the constraints are met. Our laboratory has developed and evaluated several graphical formats through which this inherently nonvisual, but spatial, information can be presented (fig. 1). '

The most basic format, developed first, is a set of horizontal bar graphs in which each of the six forces and torques is displayed (fig. 1a) (ref. 5). This type of display has been tested in our laboratory (ref. 10) and in the simulated Space Shuttle cargo bay at the Johnson Space Flight Center (ref. 9) where it has been shown to reduce the magnitude and duration of forces required to complete a task. However, the horizontal bar graph display fails to represent the spatial content of the force/torque information because the assignment of the forces and torques to the bars is essentially arbitrary.

In the JPL-OMV (Orbiting Maneuvering Vehicle) smart hand (ref. 6), an improved display was developed which represented a primitive perspective view of the unit vectors making up the hand reference frame (fig. 1b). Torques were represented by bar graphs crossing the appropriate axes. This type of display was tested in over 21 hr of experimentally recorded teleoperation with operators of various experience levels performing simulated satellite servicing tasks (ref. 11).

In these experiments, the JPL-OMV smart hand was mounted on the Prototype Flight Manipulator Arm (PFMA) at Marshall Space Flight Center, and operators performed task-board operations from a remote control room. The operators were provided with three camera views of the scene, a six-axis "joyball" controller (ref. 12), and the previously described star display of forces and torques. Task performance was measured in terms of RMS forces/torques required to perform the task. Low RMS forces/torques indicated the absence of forcing and jamming of the tool and thus better task performance. Force and torque display was available to the operators in selected trials and comparative force control performance was measured for the two cases. Although operators reported that the visual force/torque information was useful, no significant reduction was observed in task-related forces and torques.

When an earlier version of this display was tested on the space shuttle RMS simulator (ref. 13), reductions in forces were demonstrated. This discrepancy can be attributed to the relatively poor position-control performance of the PFMA and its high stiffness, versus the highly accurate position control capability of the RMS and its low stiffness. In the absence of true force-control capability, operators apparently adopt a strategy of controlling forces by commanding small position increments against the stiffness of the manipulator and load.

This type of indirect control strategy demonstrates that in the case of telemanipulation, it is very difficult to evaluate displays in isolation–especially in terms of overall task performance.

A further display refinement is to generate a three-dimensional bar graph in which the magnitude of each force component is displayed in the direction of its unit vector. Torques are displayed as circular bar graphs centered on the axes. This display has been rendered in color and true perspective on an IRIS graphics workstation (fig. 1c). Evaluation in use awaits integration of the IRIS with actual telerobotic hardware.

## Event-Driven Displays

We are currently developing enhancements to improve these force/torque displays. In many tasks, the desired outcome is to perform a manipulation subject to specific constraints. For example, the task may be to press on a latch such that z-axis force is greater than or equal to 10 lb and x and y forces and all torques are less than 1 lb (or ft-lb). The burden of checking these constraints can be removed from the operator by a set of display primitives which indicate the constraints on each axis and a global flag indicating that all constraints are satisfied. These "event-driven" displays also have served to combine information from proximity, tactile, and force/torque sensors (refs. 14 and 15).

This concept has been tested with a light-emitting diode (LED) version of the star pattern display at Johnson Space Center (ref. 5) and has been added to the OMV smart-hand display.

One key issue is the value of detailed visual force/torque information to the operator relative to other visual information sources, especially cameras. Future experimentation will address this question by forcing the operators to choose among display sources and recording relative frequency of selection of each display. A cost will be imposed for switching between sources to prevent the adoption of a scanning strategy. Thus, for example, operators will attempt to minimize a time score for completing a manipulation task, but will be penalized 1 sec each time they switch from the various displays.

In present force/torque displays, the operator must transform the display from the hand frame to the static frame for the display. The manipulator control device is usually referenced to task space and the operator can be assumed to map the various camera views to a mental model of the task space. Knowledge of the position and orientation of the manipulator end effector in task space is required to perform this mapping. Incorporating this mapping into a task-space display presents the technical issue of interfacing the hand electronics to the manipulator control system (to provide task space position and orientation), and the design issue of how to present the task-space information. Alternatives being considered are to transform the star display to the end-effector position and orientation, transform it again to the viewplane of one of the cameras, and superimpose it on the camera view. Another possibility is to create a synthetic deformable object such as a striped cylinder, locate it at the manipulator wrist, and deform it according to the forces and torques present at the wrist. A display of the deformed cylinder superimposed on the video scene would give an easy-to-grasp, intuitive picture of the manipulator's interaction with its environment.

## Real-Time Perspective Simulation

Simulation presents an effective means of developing teleoperator systems, can provide valuable feedback during the use of such a system, and can be an effective design tool.

In our laboratory setup, a universal 6 degree-of-freedom, force-reflecting hand controller (FRHC) is used as master and a PUMA 560 robot is used as slave. The kinematics and dynamics of both arms are extensively studied and described in the literature (refs. 16 and 17). Two National Semiconductor NS-32016 microprocessors were chosen to control the FRHC and the PUMA arm, respectively. The distributed control and interface information is detailed in reference 18. A real-time simulator also was built in parallel with the distributed control system to facilitate human control performance studies, hardware/software checkout, and operator training.

The real-time simulator (fig. 2) consists of almost all the hardware of the complete telemanipulation system except that the PUMA manipulator is replaced by the computer graphic simulation. The 6 degree-of-freedom FRHC is the key interface between the operator and the control station. It provides the necessary force feedback to the operator and is equipped with six optical encoders for position sensing and six motors for backdriving the operator. The control-station processor interprets the encoder values and converts them into joint angles. It then performs forward kinematics calculations to determine the end position of the FRHC in the work space and then transmits those position commands to the remote station. The remote processor receives the position command from the control station, computes inverse kinematics of the PUMA arm, and determines the desired joint angles which are sent to the graphics processor for animation. The Silicon Graphics IRIS work station is employed for the graphics generation and display. It animates the movements of the PUMA arm in color graphics and provides the task-simulation environments.

The requirements for the display are that animation be generated at a rate high enough that the simulation appears continuous and realistic to the operator. The Silicon Graphics IRIS 2400 is a UNIX-based graphics workstation which uses a highly pipelined display architecture. The IRIS 2400 contains several VLSI hardware graphics processors known as geometry engines (ref. 19). These are capable of performing basic graphics operations, such as matrix transformations, clipping and mapping to device coordinates at a rate of approximately 65,000 three-dimensional, floating-point coordinates per second. The geometry engines are arrayed to form the

graphics pipeline, with a 68000 microprocessor used as a low-level pipeline manager. The host processor for the geometry pipeline is a 68010 which communicates with it over the multibus. The geometry engines are accessible via the C graphics library provided with the system. This library enabled high-level operations such as coordinate frame and polygon definitions to be specified from within the Applications Program.

The PUMA-560 model was created into two steps. The constituent graphical objects such as the links and base were defined relative to their own coordinate frames. Appropriate coordinate frames were then developed for each link in a fashion similar to the Denavit-Hartenberg link specifications. Links were subsequently displayed in their appropriate coordinate frames, thus forming the complete model of the robot. The necessary link parameters were found in Craig (ref. 17).

The frame transformations for the forward kinematics of the PUMA were inherent to the graphical model of the PUMA. Robot motion animation was achieved by varying the appropriate link parameters, i.e., the joint angles, and rapidly redrawing the robot model according to these new values.

The capability to perform high-speed graphics computations permitted the display of a model of intermediate complexity at approximately a 10-Hz refresh rate, including input/output operations. Data are sent from the hand controller to the IRIS in 16-bit binary form over the RS232 serial interface.

Hidden-surface elimination was investigated, but not implemented in this version of the simulation display because of speed constraints. Several fast software algorithms for hidden-surface elimination exist. The general principle involved is to presort the polygons composing a static object before they are displayed. Unfortunately, while these techniques work well for a roving viewpoint and static objects, the links in the PUMA model are constantly changing their position relative to each other and thus their constituent polygons are not presortable.

There are many applications for the graphics simulation of the PUMA 560 running on the IRIS workstation. Of most value is its use as a debugging tool. Many software modules developed for control of the manipulator can be tested and debugged using the graphics simulation without actually using the manipulator. In general, the simulation allows its users to test-control software when the actual manipulator is not available, or its design is not yet finalized. Different manipulator geometries can be explored for functionality before they are actually prototyped. This flexibility is true for hand-controller design as well.

The simulation also can be used as a tool for training teleoperator system operators. Fictitious objects can be introduced into the virtual work environment so that operators can practice pick-and-place tasks as well as more complex operations without endangering hardware. Using the IRIS system's ability to clip against an arbitrary plane, end-effector collision with objects in the virtual environment could be detected and indicated in real time. This feature will assist the operator in practicing teleoperation and collision avoidance.

When a significant time delay in communication exists between the controller and manipulator (e.g., Earth-based control station commanding a geosynchronous satellite servicing teleoperator) a graphic simulation could become valuable in enhancing operator performance. By overlaying a stereoscopic wire-frame view of the manipulator on the stereoscopic television image of the task space, a predictive display can be obtained (ref. 20). This allows the operator to immediately

realize the ramifications of his/her actions before actually performing an operation. Commands could be buffered and then sent once the operator is sure that no damage will result from given actions.

## Teleoperator Laboratory Design Simulation

We have also used perspective displays as a design tool to explore various layouts for a dual-arm telemanipulation laboratory. These simulations (fig. 3a) allowed the designers to specify robot base location and posture (elbow up/down, shoulder in/out, etc.) in a model of the actual laboratory space. A grid placed on the floor represents the actual floor tiles so that the simulation can be easily related to the actual space. A projection of the maximum extent of each robot's work volume was drawn on the floor grid. The intersection of the two work-volume projections gives an idea of the cooperative work volume of the two robots. Note that the work volume is a function of arm configuration if arm flips are not allowed. Another important design issue directly addressed by this display is the visibility of the task space and especially the manipulator end effectors by the operator (in direct operation from the control station) or from a particular camera. Because the viewpoint of the simulation can be changed dynamically, designers can view the robots from any contemplated control station or camera mount. On the basis of this simulation, the plan shown in figure 3 was shown to have higher cooperative work volume and better sight lines from the operator control station than a competing plan.

## Simulated Satellite Servicing Animation

Autonomous task-sequence simulation takes the static scene simulation a step further by adding the element of time and order of subtask execution. Our application is an animation of two robots performing the replacement of an attitude-control system on the Solar Max Satellite (fig. 3b). This is the chosen scenario for the 1988 telerobot demonstrator project at JPL. The simulation is adaptable to a variety of tasks, and could take input from artificial intelligence task planners to provide a means of human verification of the output of autonomous subsystems.

## Executive Control Displays

A complete telemanipulation system requires far more interaction with operators than that required for the purely manipulation components of a task. Considerable human interaction overhead will be required to control cameras, select system operating mode, attend to error conditions, start up and shut down the system, and hand off control between teleoperation and automatic operation. Many of today's telemanipulation systems require a second operator and control station to perform these "executive" functions. The nature of this task is to selectively attend to details of whichever one of a large number of subsystems requires attention.

## Desktop Control Station

The traditional approach to this executive control station is a console or series of racks filled with a separate control panel for each subsystem. An attractive alternative is offered by a single

control station consisting of a large bit-mapped display through which an operator can control all of these functions.

We are currently prototyping such an executive-control display which compresses all executive-control functions into a single high-resolution workstation screen (fig. 4). Control interaction will take place between the workstation and the subsystems over a local area network.

The basis for this display is the desktop direct-manipulation environment (refs. 21, 22, and 23) as implemented in the Macintosh and the Sun workstation, and pioneered in the Smalltalk environment (ref. 24), which evolved from earlier work such as Sutherland's Sketchpad system (ref. 25). The workstation screen represents a domain which is populated with icons representing the various systems. The operator can expand a subsystem icon to reveal a complete control panel for that system containing buttons, indicators, sliders, graphics displays, and so forth. Icons can be dynamic so that alarm conditions can reach the attention of the operator when a subsystem is closed.

We have prototyped examples of icons from such a system on a Sun workstation (at the display and human-interaction level only). A control station based on this concept will take up much less space than a conventional panel rack and will be very flexible with respect to future expansion. Operators could easily customize the display to the requirements of a specific task.

Interaction between the manipulation operator and the icon-based executive control station is desirable, eliminating the need for a second operator even in two-handed teleoperation. In currently planned dual-arm teleoperation systems, the operator's hands are occupied with controlling two slave manipulators through six-axis, force-reflecting, hand controllers. Either hand controller (depending on operator preference) could be temporarily changed over to controlling a display cursor on the executive-control station.

## Hand Controller as Mouse

In this concept, the operator will press a button mounted on the hand controller, which will lock the slave manipulator, or turn its control over to an automatic or intelligent control system. Two degrees of freedom of the hand controller would then control the location of the cursor on the executive control display. The hand-controller backdrive capability could be used for providing detents indicating cursor position relative to the icons. This will provide an active assist in moving the cursor to small icons or panel objects. Designation of display objects (analogous to clicking a mouse button) will be accomplished by the hand-controller button normally used for gripper control. Other hand-controller degrees of freedom could be used to operate panel items such as knobs.

For example, to adjust an analog quantity such as a rate limit, the operator could move the hand controller and thus a screen cursor to a picture of a knob representing the appropriate quantity. The location of the cursor on the screen will be taken from the x and y coordinates of the hand controller. In the immediate neighborhood of the "knob," the operator will feel a small force generated by the control computer to represent the negative gradient of a small "potential function" on the workstation surface. The potential function contains "wells" around each of the icons and panel items. This force will guide the operator to the icon and correct small positioning errors. When the cursor is over the knob, the roll axis of the hand controller would be used to change its setting. Other types of icons would be operated by orthogonal hand-controller motions. For example, a

toggle-switch icon would be operated by the pitch axis. This provides a measure of safety because each icon can be activated only by a particular hand motion. The icons should be designed and linked to the axes of hand motion so the way to actuate them is intuitive.

At the conclusion of the executive control function, the operator would resynchronize the slave manipulator with the master and resume manipulation. The details of the transitions between manipulation control and cursor control are complex, but identical in principle to those used in the indexing function already designed into such systems.

# REFERENCES

1.  R. Miller, M. Minsky, D. Smith, and D. Akin, Space Applications of Automation and Robotics and Machine Intelligence Systems (ARAMIS), MIT, NASA Marshall Space Flight Center, Aug. 1982.

2.  W. S. Kim, S. R. Ellis, M. E. Tyler, B. Hannaford, and L. Stark, "Quantitative Evaluation of Perspective and Stereoscopic Displays in Three Axis Manual Tracking Tasks," IEEE Trans. Systems Man & Cybernetics, vol. SMC-17, no. 1, pp. 61-72, 1987.

3.  L. Stark, W. S. Kim, F. Tendick, B. Hannaford, S. Ellis, et al., "Telerobotics: Display, Control, and Communication problems," IEEE J. Robotics Automation, vol. RA-3, pp. 67-74, 1987.

4.  A. K. Bejczy, G. Bekey, and S. K. Lee, "Computer Control of Space Borne Teleoperators with Sensory Feedback," Proc. IEEE Conf. Robotics and Automation, pp. 205-214, 1985.

5.  A. K. Bejczy and R. S. Dotson, "A Force Torque Sensing and Display System for Large Robot Arms," Proc. IEEE Southeastcon82, Destin, FL, April 1982.

6.  A. K. Bejczy and B. Jau, "Servicing with Smart End Effector on OMV Manipulator," Proc.: Satellite Services Workshop II, NASA Johnson Space Center, Nov. 1985.

7.  J. E. Pennington, A Rate Controlled Teleoperator Task with Simulated Transport Delays, NASA Langley Research Center, Sept. 1983.

8.  M. Salganicoff, E. Austin, and C. Fong, "Real-time Simulation of a Distributed Teleoperator," Proc. Intl. Assoc. Sci. Technol. for Development, Vancouver, June 1986.

9.  A. K. Bejczy, R. S. Dotson, J. W. Brown, and J. L. Lewis, "Force-Torque Control Experiments with the Simulated Shuttle Manipulator in Manual Control Mode," Proc. 18th Ann. Conf. on Manual Contr., Dayton, OH, June 1982.

10. A. K. Bejczy and M. Handlykken, "Experimental Results with a Six-Degree-of-Freedom Force Reflecting Hand Controller," Proc. 17th Ann. Conf. Manual Contr., Los Angeles, June 1981.

11. B. Hannaford, "Task Level Testing of the JPL-OMV Smart End Effector," Proc. JPL-NASA Workshop on Space Telerobotics, Pasadena, CA, Jan., 1987.

12. A. L. Lippay, M. King, and R. V. Kruk, "Helicopter Flight Control with One Hand," Can. Aeronaut. Space J., vol. 31, no. 4, pp. 225-345, 1985.

13. K. Corker, A. K. Bejczy, and B. Rapparport, "Force/Torque Display for Space Teleoperation Control Experiments and Evaluation," Proc. 21st Ann. Conf. Manual Contr., Columbus, OH, June 1985.

14. A. K. Bejczy and G. Paine, "Event-Driven Displays for Manipulator Control," Proc. 14th Ann. Conf. Manual Contr., 1978.

15. G. Paine and A. K. Bejczy, "Extended Event-Driven Displays for Manipulator Control," Proc. 15th Ann. Conf. Manual Contr., Dayton OH, March 1979.

16. R. P. Paul, in Robot Manipulators–Mathematics, Programming, and Control, Boston, MIT Press, 1981.

17. J. Craig, Introduction to Robotics: Mechanics and Control, Addison-Wesley, 1986.

18. C. P. Fong, A. K. Bejczy, and R. Dotson, "Distributed Microcomputer Control System for Advanced Teleoperators," Proc. IEEE Intern. Conf. Robotics and Automation, San Francisco, April 1986.

19. J. H. Clark, "The Geometry Engine: A VLSI Geometry System for Graphics," Computer Graphics, vol. 16, no. 3, July 1982.

20. T. B. Sheridan and W. R. Ferrell, "Remote Manipulative Control with Transmission Delay," IEEE Trans. Human Factors in Electronics, vol. HFE-4, pp. 25-29, 1963.

21. B. Shneiderman, "Direct Manipulation: A Step Beyond Programming Languages," IEEE Computer, vol. 16, p. 8, 57-69, Aug. 1983.

22. B. Shneiderman, Designing the User Interface: Strategies for Effective Human-Computer Interaction, Addison-Wesley, 1987.

23. H. J. Bullinger, K. P. Fahnrich, and J. Zeigler, "Human-Computer Interaction and Direct Manipulation," in HUFIT Consortium Project Report, Fraunhor Institute IAOP FhG, Stuttgart, 1984.

24. A. Goldberg, Smalltalk-80: The Interactive Programming Environment, Addison-Wesley, 1984.

25. I. E. Sutherland, "Sketchpad: A Man-Machine Graphical Communication System," AFIPS Spring Joint Computer Conf., vol. 23, pp. 329-346, 1963.
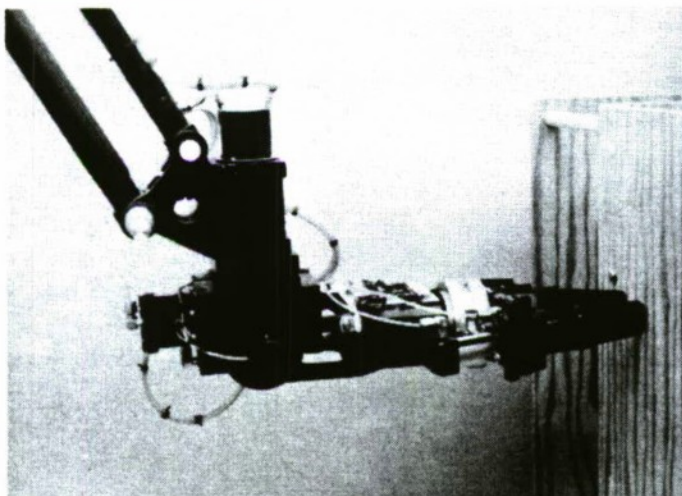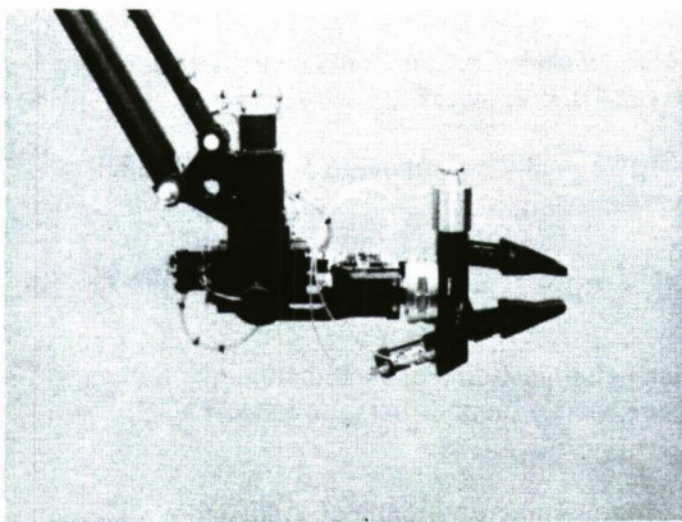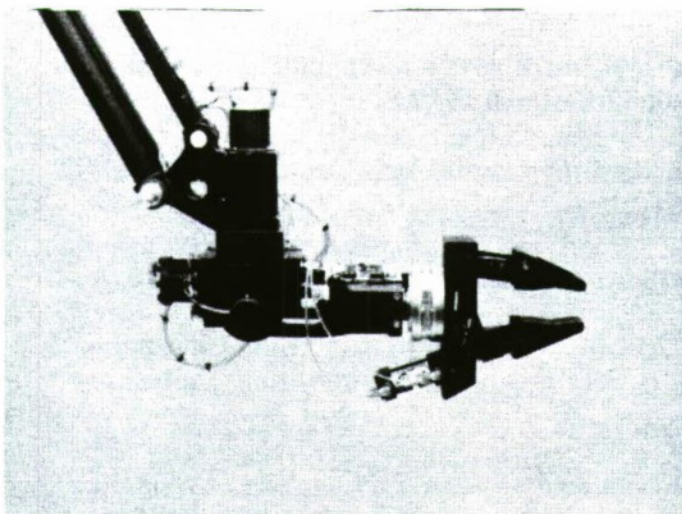
(a) A simple set of horizontal bar graphs, one bar for each of the six axes of force and torque.

Figure 1.– Displays of force/torque information for telerobotics. Several formats have been developed and experimentally evaluated at JPL for the display of forces and torques encountered by a remote manipulator to the controlling operator. Panels (a) and (b) represent monochrome displays, (c) represents color.
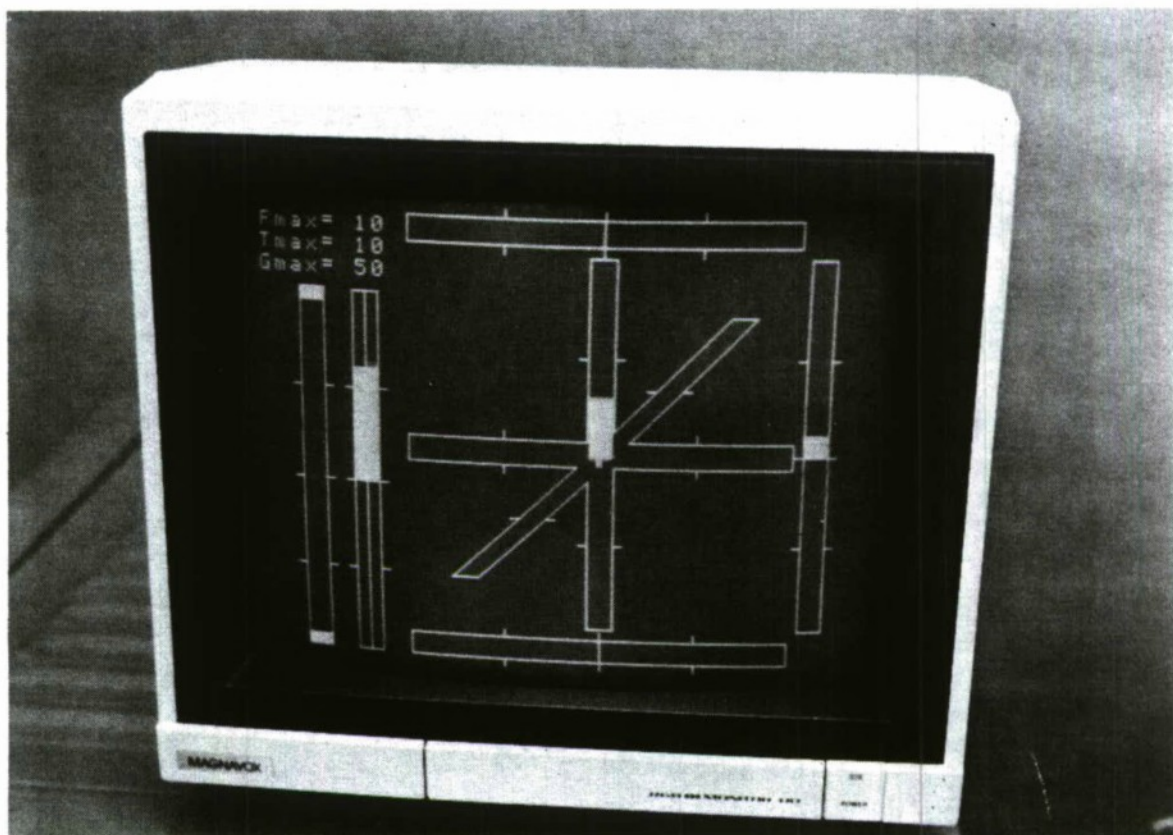
(b) A pseudo-perspective display in which the bar graphs are aligned with unit vectors representing the direction of action of forces, and the roll, pitch, and yaw axes for torques.

Figure 1.– Continued.
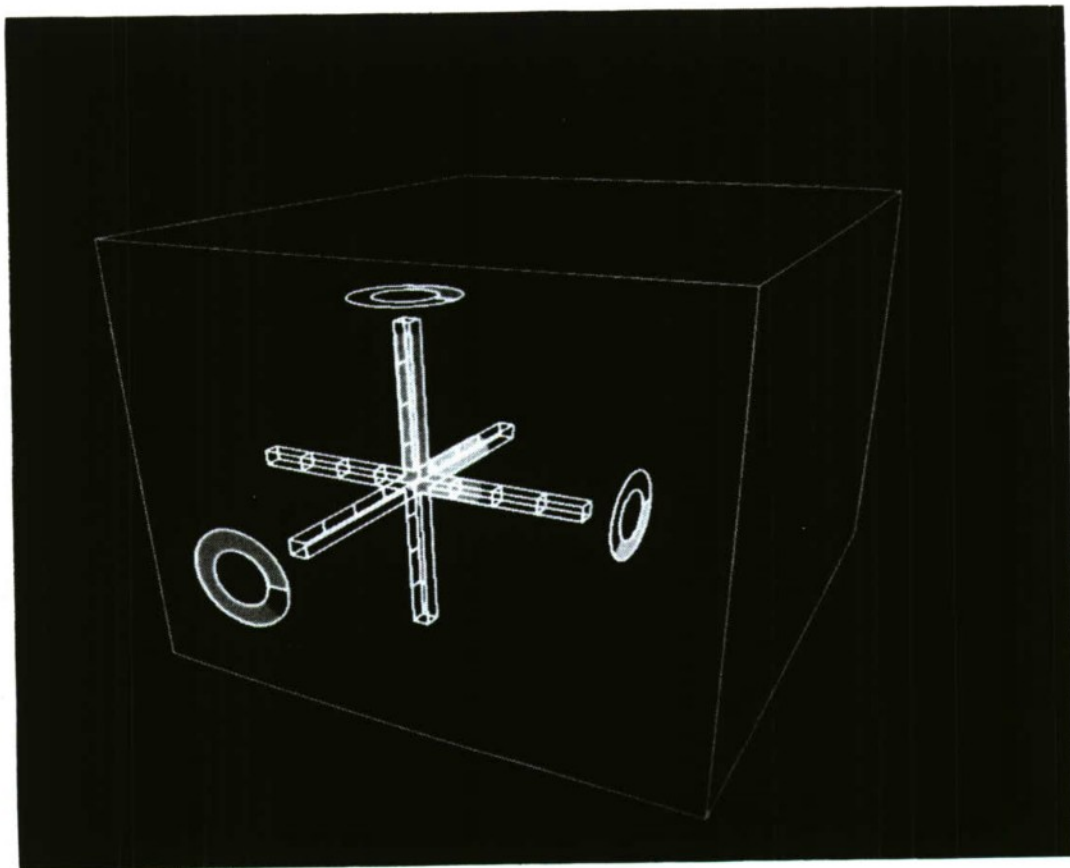
c) A true perspective, full-color display.

Figure 1.– Concluded.

Figure 2.– Real-time simulation of a robot manipulator in telemanipulation. The wire-frame simulation substitutes for the manipulator for software validation or operator training. The complete telemanipulation system consists of a hand controller (left); control processors (not shown); monochrome display; and optionally, robot manipulator (background). The display computer is plug-compatible with the manipulator controller.

(a)
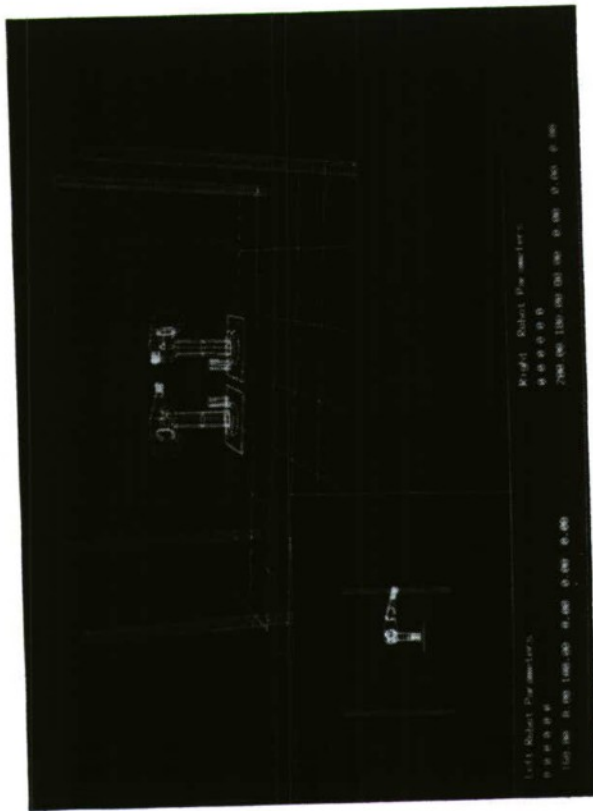


(a)
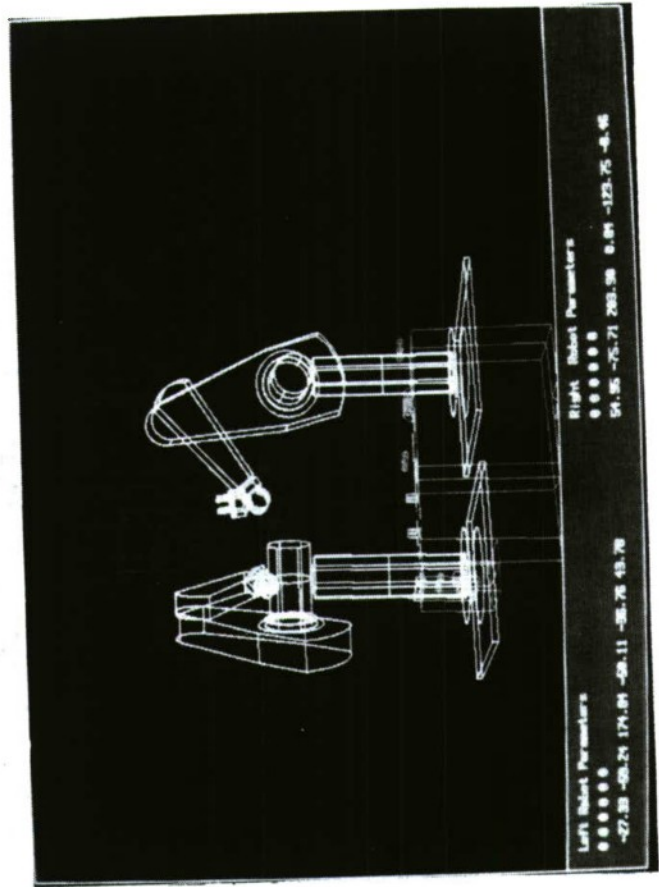


(b)

Figure 3.– Design and analysis simulation displays. Graphics simulation has been successful in the design of the kinematics and visual aspects of telemanipulation systems. In the dual-arm laboratory simulation (a) the designers can specify manipulator base locations and joint angles, and then dynamically rotate the display viewpoint (or select the orthogonal projection direction) to assess cooperative work volume (note projections on floor) and sight lines. Shown in (a) are two displays of the same equipment configuration. The right panel shows the view from the control station window. In the satellite-servicing animation (b), the complete sequence of operations involved in two PUMA arms replacing a module on the Solar Max Satellite is simulated. (a,b,c are color displays.).
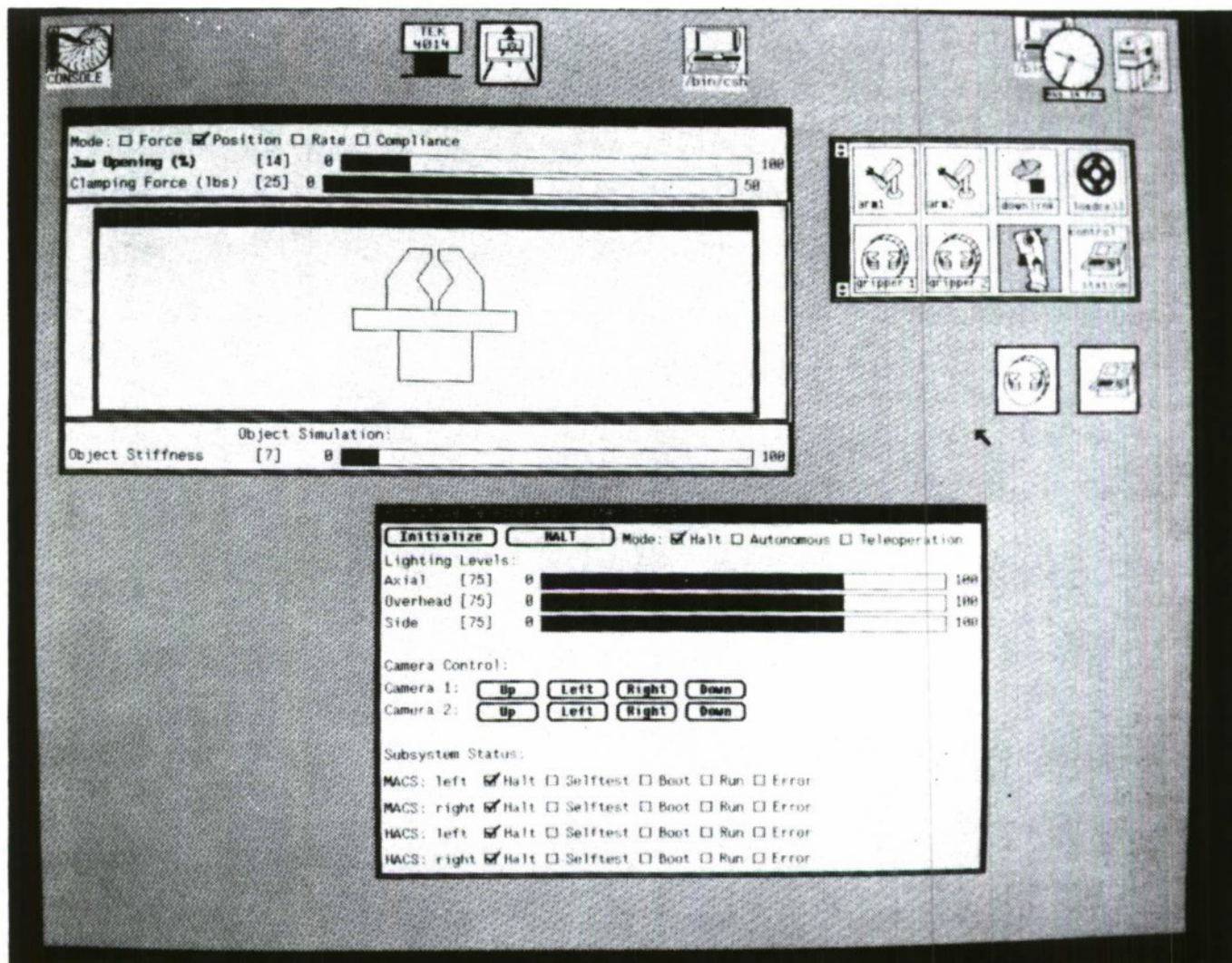
Figure 4.– Executive-control display for a telemanipulation system. Icons representing each of the many subsystems involved in a full, dual-arm, telemanipulation system are displayed on a single monochrome workstation screen. Subsystems are controlled through a pointing device operating simulated switches, sliders, and buttons. In conventional systems, these functions are controlled by a second operator sitting at a large rack of hardware control panels. The executive control display can eliminate the need for a second operator because the manipulation operator can operate the display using the force-reflecting hand controller.

# PERCEPTION-ACTION RELATIONSHIPS RECONSIDERED IN LIGHT OF SPATIAL DISPLAY INSTRUMENTS

Wayne L. Shebilske
Department of Psychology
Texas A&M University
College Station, Texas

## SUMMARY

Spatial display instruments convey information about both the identity and the location of objects in order to assist surgeons, astronauts, pilots, blind individuals, and others in identification, remote manipulations, navigation, and obstacle avoidance. Scientists believe that these instruments have not reached their full potential and that progress toward new applications, including the possibility of restoring sight to the blind, will be accelerated by advancing our understanding of perceptual processes. This stimulating challenge to basic researchers was advanced by Paul Bach-Y-Rita (1972) and by the National Academy of Science (1986) report on Electronic Aids for the Blind. Although progress has been made, new applications of spatial display instruments in medicine, space, aviation, and rehabilitation await improved theoretical and empirical foundations.

## GAPS IN OUR UNDERSTANDING OF PERCEPTION-ACTION RELATIONSHIPS

What is it that applied researchers want to know that basic researchers can't tell them?

Inadequacies of the present foundations are revealed by considering a discrepancy between issues that are addressed by basic researchers in the field of perception and questions that are asked by developers of spatial display instruments. These groups have different perspectives on two major functions of our sensory system, which are 1) to provide a conscious representation of spatial-temporal relationships, and 2) to guide our performance as we interact with our environment. Perception researchers concentrate on the first of these functions, providing perceptual impressions (subjective experiences) of objects or events such as apparent shape, size, orientation, and movement. They describe how the world does appear to us, and they analyze the determinants of our subjective experience of the world. In contrast, human factors engineers, clinicians, and specialists in artificial intelligence develop spatial display instruments to enhance performance that depends upon sensory information. Consequently, they ask questions about the second function of the sensory system, guiding performance. Thus, there is a gap between the main issues that are addressed by researchers in the field of perception and the information that is needed by developers of spatial display instruments.

Ironically, this gap has gone unattended because a corresponding gap has existed for a long time in researchers' understanding of the relationships between stimulus information, perceptual impressions, and performance. One major approach to research, the direct perception approach, bases its research on the untested assumption of a one-to-one correspondence between stimulus

information and performance (e.g., Gibson, 1979; Turvey and Solomon, 1983). Other major approaches, mediated perception approaches, base their research on untested assumptions about the relationship between appearance and performance. Experimental tasks depend upon the availability of a representation of spatial-temporal relationships, and it is often assumed that the representation upon which performance is based corresponds to perceptual impressions of spatial-temporal relationships. Some paradigms carry this untested assumption to an extreme by inferring registered values of space in one task from perceptual impressions on a different task (e.g., Gogel, 1980). Accordingly, both direct perception researchers and mediated perception researchers have substituted untested assumptions for an empirically based theoretical foundation for understanding relationships between stimulus information, perceptual impressions, and performance.

Previous literature suggests that these relationships are complex and variable from situation to situation. During natural events in information-rich environments, there sometimes is a one-to-one correspondence between stimulus information and performance (e.g., Lee and Reddish, 1981; Turvey and Carello, 1986; Warren, 1984) and there sometimes is not (e.g., Shebilske, 1981, 1987a, 1987b; Shebilske, Karmiohl, and Proffitt, 1984). This variability is complicated by the fact that there is no general way to predict what the relationship will be in any given natural event.

Understanding the relationship between perceptual impressions and performance is similarly complicated not only by evidence that there are at least three modes of perceptual impressions (Rock, 1983) and that instructions can affect which one of these modes will correlate with performance (e.g., Carlson, 1977; Leibowitz and Harvey, 1969; Ebenholtz and Shebilske, 1973), but also by the finding of both tight and loose relationships. At one extreme, there is evidence for a very tight relationship (e.g., Coren, 1981). At the other extreme, there is evidence of very loose relationships. Examples include subliminal priming, which is an "unseen" word facilitating the recognition of another word (Marcel, 1983); blindsight, which is pointing at targets that cannot be "seen" (Bridgeman and Staggs, 1982); and paradoxical perceptions, such as apparent motion without apparent change in position (Shebilske and Proffitt, 1983).

Attempts to explain this variability include arguments for top-down influences. For example, Gogel (1977) stated that objects can be cognitively judged to be in a different location than they appear and that performance can reflect these cognitive judgments. Bottom-up influences have also been proposed. Shebilske and Proffitt (1981) suggested, for example, that apparent motions of a stimulus during head movements might be based "solely on motion information and principles of perceptual organization that make no use of distance information." Simultaneously, the same stimulus might elicit pointing responses that are based on distance information from one set of sources and size estimations that are based on distance information from another set of sources.

The problem is that our empirical and theoretical foundation is inadequate to predict when top-down and/or bottom-up influences will alter the relationships between stimulus information, perceptual impressions, and performance. The consequence of this inadequacy is, at the very least, a bottleneck in the transfer of information from basic research about perception to applications that depend upon sensory input, such as spatial display instrumentation. An even worse consequence is the danger of undermining parts of our basic research foundation that are based upon untested assumptions about these relationships.

# ECOLOGICAL EFFERENCE MEDIATION THEORY

Operations for encoding sensory information should approach optimal efficiency in the environment in which a species evolved, according to an ecological point of view (Gibson, 1979; Shebilske and Fisher, 1984; Shebilske, Proffitt, and Fisher, 1984; Turvey, 1979; Turvey and Solomon, 1983). Based on this axiom and the observation that efference-based and higher order light-based information interact to determine performance during natural events, Shebilske (1984, 1987a, 1987b) proposed an Ecological Efference Mediation Theory of natural event perception. According to this theory, both the phylogeny and the ontogony of the visual system are shaped by internal state variables as well as by environmental variables. When the preceding discussion is recast in terms of this theory, the question becomes: How can fluctuations in relationships between stimulus information, perceptual impressions, and performance afford an adaptive advantage relative to all the conditions to which humans are exposed? Attempts to answer this question resulted in a hypothesis about Ecologically Insulated Event Input Operations (EIEIO). This EIEIO hypothesis will be explained in the remainder of this essay.

## The EIEIO Hypothesis

Humans are able to perform in a wide range of transient internal and external states. The EIEIO hypothesis accounts for this flexibility by postulating separate input modules that are molded by interactions of an organism with its environment in an attempt to achieve maximally efficient performance of sensory guided skills within the prevailing internal and external states in which the skill is performed. Schmidt (1987) reviewed the history of thought on the theme that practice can change the way sensory information about the world is used to guide performance. He started with William James' observation (1890) that practice of skills seems to lead to more automatic, less mentally taxing behavior. This observation spawned considerable research leading to evidence for three separate process level changes that seem to contribute to this practice effect as follows: 1) tasks that are slow and guided shift from dependence on exproprioceptive information to dependence on proprioceptive information (e.g., Adams and Goetz, 1973); 2) tasks that have predictable parameters, such as predictable target locations in pointing tasks, shift to open-loop control (e.g., Schmidt and McCabe, 1976); and 3) tasks that have unpredictable parameters shift to fast, automatic, and parallel processing of the information needed to make decisions (e.g., Schneider and Shiffrin, 1977; Shiffrin and Schneider, 1977). The EIEIO hypothesis is a proposal of a fourth manner in which practice can change the way sensory information is used to guide performance. The proposal is that the bases for sensory guided performance can shift from conscious representations of spatial-temporal relationships to EIEIO representations that do not correspond to conscious perceptual impressions. In contrast to the other three mechanisms, which were identified through studies contrasting variables that are an integral part of the task, the EIEIO hypothesis emerged from considerations of the various internal and external contexts in which skills are performed. The EIEIO hypothesis encompasses five testable premises.

Premise 1. In addition to performance being guided by representations that correspond to conscious perceptual impressions of spatial-temporal relationships, performance can also be guided by one or more abstract, symbolic EIEIO representations of the same spatial-temporal relationships.

Premise 2. These EIEIO representations are insulated from each other and from the conscious one in the sense that they can be altered independently.

Premise 3. Differences between the accuracy, speed, and attention demands of EIEIO representations result from: 1) separate selective attention mechanisms that result in the picking up and processing of different potential sources of information, 2) different parsing routines that result in sampling units of different spatial sizes and/or different temporal durations, 3) different weightings that are used for various sources of information, and 4) different rules (e.g., rigidity assumption) and/or different principles of processing (e.g., minimum principle) that are used.

Premise 4. Conditions leading to the development and use of EIEIO representations during phylogeny or ontogeny depend upon interactions between an organism and its environment. Modules for forming EIEIO representations will result when an organism has the opportunity to perform the same skill repeatedly in an environment that 1) has contextual variability over a range that is narrower than the entire range in which the more general system must operate and 2) provides an opportunity to learn that the conscious representation is less efficient than an alternative one. The EIEIO representations that develop are utilized only when a skill is performed in the environment in which it was learned.

Premise 5. Whereas input operations corresponding to conscious representations are designed to be maximally efficient over the entire range of contextual variability to which an organism is exposed in its environmental niche, EIEIOs are designed to be maximally efficient within a narrower range of contextual variability within which a particular skill is performed. This premise is related to a familiar design for adaptability in biological systems. It is common to have a relatively narrow range of sensitivity available at any one moment, but to have this narrow range move over a much broader range in order to adapt to prevailing conditions. An example is light and dark adaptation in which a relatively narrow range of sensitivity to light exists at any given moment. But the absolute level of this momentary range can be adjusted up (light adaptation) or down (dark adaptation). The proposed design of EIEIOs, however, has an important unique feature. Specifically, a conscious representation that is based on very generalizable input operations is always available during normal waking consciousness as long as the stimulus information is above the momentary sensory threshold (or signal-detection criterion). However, after an extended opportunity to perform a skill under conditions that consistently have a relatively narrow range of contextual variability of internal and external states, the function of the conscious representation in guiding performance on that specific skill can be momentarily replaced by EIEIO representations that are more efficient within the prevailing narrow range of contextual variability. For example, gymnasts might be able to form more efficient EIEIO representations to guide their skilled performance by having their input operations take advantage of the fact that their skill is always performed in a well-lighted, highly structured environment. At the same time, the gymnasts would retain the more generalizable input operations that would result in continual access to a conscious representation at all times during normal waking consciousness, including whenever the gymnasts darted in and out of all the environmental conditions with which humans can be confronted.

## CONCLUSIONS

Progress toward realizing the full potential of spatial display instruments is limited less by technology than by an inadequate understanding of perceptual processes. A bottleneck is

encountered in understanding the relationships between stimulus information, experiential responses, and performance. In previous articles, I have taken stands against the postulation of a one-to-one correspondence in these relationships, and I have argued against development of theories, research methodologies, and applications based on this postulation. Here, I argued for steps aimed at developing a theoretical and empirical foundation for understanding, predicting, and controlling the perception-action link.

I reviewed three ways that have been proposed for how perception-action relationships can change. I then proffer a fourth way, the EIEIO hypothesis, which included five testable premises about the impact of contextual variability on perception and performance. Testing these premises in contexts that are relevant to spatial display instruments will advance spatial instrument technology by enhancing our ability to understand, predict, and control the many-to-one correspondence that often exists between stimulus information, perceptual impressions, and performance.

# REFERENCES

Adams, J. A., and Goetz, E. T.  Feedback and practice as variables in error detection and correction.  J. Motor Behav., 1973, 5, 217-224.

Bach-Y-Rita, P.  Brain mechanisms in sensory substitution, New York:  Academic Press, 1972.

Bridgeman, B., and Staggs, D.  Plasticity in human blindsight.  Vision Res., 1982, 22, 1199-1203.

Carlson, V. R.  Instructions and perceptual constancy judgments.  In W. Epstein (Ed.) Stability and constancy in visual perception:  mechanisms and processes, New York:  Wiley, 1977.

Committee on Vision, National Research Council.  Electronic travel aids:  new directions for research.  Washington, DC:  National Academy Press, 1986.

Coren, S.  The interaction between eye movements and visual illusions.  In D. F. Fisher, R. A. Monty, and J. W. Senders (Eds.) Eye movements:  cognition and visual perception.  Hillsdale, NJ:  Erlbaum, 1981.

Ebenholtz, S. M. and Shebilske, W. L.  Instructions and the A and E effect in judgments of the vertical.  Amer. J. Psychol., 1973, 86, 601-612.

Gibson, J. J.  The ecological approach to visual perception.  Boston:  Houghton-Mifflin, 1979.

Gogel, W. C.  The metric of visual space.  In W. Epstein (Ed.) Stability and constancy in visual perception:  mechanisms and processes, New York:  Wiley, 1977.

Gogel, W. C.  The sensing of retinal motion.  Percept. Psychophys., 1980, 28, 153-163.

James, W.  The principles of psychology (Vol. 1), New York:  Holt, 1890.

Lee, D. N. and Reddish, P. E.  Plummeting gannets:  a paradigm of ecological optics.  Nature, 1981, 293, 293-294.

Liebowitz, H. W. and Harvey, L. O., Jr.  Effect of instructions, environment, and type of test object on matched size.  J. Exp. Psychol. 1969, 81, 36-43.

Marcel, A. J.  Conscious and unconscious perception:  An approach to the relationship between phenomenal experiences and perceptual processes.  Cognitive Psychol., 1983, 15, 238-300.

Rock, I.  The logic of perception.  Cambridge:  MIT press, 1983.

Schmidt, R. A.  The acquisition of skill:  some modifications to the perception-action relationship through practice.  In H. Heuer and A. F. Sanders (Eds.), Perspectives on perception and action, Hillsdale, NJ:  Erlbaum, 1987.

Schmidt, R. A. and McCabe, J. F. Motor program utilization over extended practice. J. Human Movement Studies, 1976, 2, 239-247.

Schneider, W., and Shiffrin, R. Controlled and automatic human information processing: I. Detection, search, and attention. Psychol. Rev., 1977, 84, 1-66.

Shebilske, W. L. Visual direction illusions in everyday situations: implications for sensorimotor and ecological theories. In D. F. Fisher, R. A. Monty, and J. W. Senders (Eds.) Eye movements: cognition and visual perception. Hillsdale, NJ: Erlbaum, 1981.

Shebilske, W. L. Context effects and efferent factors in perception and cognition. In W. Prinz and A. F. Sanders (Eds.), Cognition and motor processes. Berlin Heidelberg: Springer-Verlag, 1984.

Shebilske, W. L. An ecological efference mediation theory of natural event perception. In H. Heuer and A. F. Sanders (Eds.), Perspectives on perception and action, Hillsdale, NJ: Erlbaum, 1987a.

Shebilske, W. L. Baseball batters support an ecological efference mediation theory of natural event perception. In D. G. Bouwhuis, B. Bridgeman, D. A. Owens, W. L. Shebilske, and P. Wolff (Eds.), Sensorimotor interactions in space perception and action, Amsterdam: North-Holland, 1987b.

Shebilske, W. L. and Fisher, S. K. Ubiquity of efferent factors in space perception. In J. L. Semmlow and W. Welkowitz (Eds.) Frontiers of engineering and computing in health care. New York: IEEE Publishing, 1984.

Shebilske, W. L., Karmiohl, C. M., and Proffitt, D. R. Induced esophoric shifts in eye convergence and illusory distance in reduced and structured viewing conditions. J. Exp. Psychol.: Human Percept. Perform., 1984, 9, 270-277.

Shebilske, W. L., and Proffitt, D. R. The priority of perceived distance for perceiving motion has not been demonstrated: critical comments on Gogel's "The sensing of retinal motion." Percept. Psychophys., 1981, 29, 170-172.

Shebilske, W. L. and Proffitt, D. R. Paradoxical retinal motion during head movements: apparent motion without equivalent apparent displacement. Percept. Psychophys., 1983, 34, 467-481.

Shebilske, W. L., Proffitt, D. R., and Fisher, S. K. Efferent factors in natural event perception can be rationalized and verified: a reply to Turvey and Solomon. J. Exp. Psychol.: Human Percept. Perform., 1984, 10, 455-460.

Shiffrin, R. M. and Schneider, W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychol. Rev., 1977, 84, 127-190.

Turvey, M. T. The thesis of efference-mediation of vision cannot be rationalized. Behav. Brain Sci., 1979, 2, 81-83.

Turvey, M. T., and Carello, C. The ecological approach to perceiving-acting: a pictorial essay. Acta Psychologica, 1986, 63, 133-155.

Turvey, M. T., and Solomon, J. Visually perceiving distance: a comment on Shebilske, Karmiohl, and Proffitt (1983). J. Exp. Psychol.: Human Percept. Perform., 1983, 10, 449-454.

Warren, W. H. Perceiving affordances: the visual guidance of stair climbing. J. Exp. Psychol.: Human Percept. Perform., 1984, 10, 683-703.

# A COMMENTARY ON PERCEPTION–ACTION RELATIONSHIPS IN SPATIAL DISPLAY INSTRUMENTS

Wayne L. Shebilske
Department of Psychology
Texas A&M University
College Station, Texas

## SUMMARY

My presentation at the conference was based on a paper that was prepared in advance and submitted for publication in this volume. In addition, the presentation included several ideas that emerged during the conference as a result of interactions with other participants. I would like to convey those ideas here along with other thoughts that occurred to me later. I will organize this commentary around three objectives: (1) to promote transfer of information across disciplines; (2) to caution basic and applied researchers about the danger of assuming simple relationships between stimulus information, perceptual impressions, and performance including pattern recognition and sensorimotor skills; and (3) to develop a theoretical and empirical foundation for predicting those relationships.

## INFORMATION TRANSFER ACROSS DISCIPLINES

This conference clearly indicated that basic and applied researchers have crossed traditional boundaries to work together toward new applications of spatial display instruments. For example, on the one hand, leaders in basic research on perception, such as Richard Gregory and Richard Held, spoke about their current research concerning applications of spatial display instruments. On the other hand, M. W. McGreevy, a leader in promoting the application of spatial display in space, also promoted basic research on sensation and perception. Thus, in place of the bottlenecks of which I spoke in my paper, I got an impression of open communication and a steady flow of information. As a result, multidisciplinary research teams have exciting agendas for research on general principles that have direct relevance to spatial display technology.

I also discovered tremendous interest in transferring information between those who are developing spatial display instruments to enhance normal sensory function or to extend it to remote-control situations and those who are developing electronic aids for the blind. I discussed with many participants of the present conference a study on the latter topic that was organized while I was Study Director for the committee on Vision (COVIS) of the National Academy of Sciences. That Committee has recently released a study on electronic aids for the blind that includes a research agenda that is highly relevant to the research programs of many of those who participated in the present conference. For example, the report calls for more research on the nature of information that is picked up about surfaces, and we saw in the present conference that this issue is also important in teleoperation of land vehicles (see McGovern, this volume). The COVIS report can be ordered by calling (202) 334-2565. You might also want to request information on a recent COVIS conference on visual displays.

# DANGERS OF ASSUMING SIMPLE RELATIONSHIPS BETWEEN PERCEPTUAL IMPRESSIONS AND PERFORMANCE: DON'T TRUST YOUR INTUITIONS

My paper reviewed evidence that relationships between stimulus information, perceptual impressions, and performance is complex, variable, and currently unpredictable. It is tempting to treat the evidence as quirks since that would make life so much easier for basic and applied researchers. If these relationships were simple, constant, and predictable, consider how worry-free one could be in making inferences about basic principles of perception from observations about performance, or in making decisions about human factors of performance from data about perceptual impressions. Several considerations add to the temptation to regard the evidence as quirks. For one thing, much of it comes from exotic clinical or laboratory situations regarding blind sight, subliminal priming of recognition, and paradoxical perceptions. Furthermore, our intuitions tell us that our sensory-guided performance corresponds to our perceptions most of the time.

With these considerations in mind, my presentation included a simple demonstration of discordance between perceptual impressions and performance in an everyday situation. I placed a plastic golf ball on a carry-out lid on an old McDonald's coffee cup and asked people to observe the ball with one or two eyes. The ball and cup were placed on an edge of a table while observers stood leaning over the cup and judged the apparent viewing distance between themselves and the ball. In agreement with data reviewed by Stanley Roscoe (this volume), participants at the conference and undergraduates tested at Texas A&M University saw the ball as being the same distance or slightly farther away (an average of about 1 cm) with one eye in comparison to the apparent distance with binocular viewing. The same observers were also asked to hit the ball off the cup by swinging a ruler parallel to the cup surface at the level of the ball. Order of these tasks was counterbalanced across subjects and the results were the same both groups. Almost all subjects swung well above the ball (an average of about 3 cm). I call the results of this demonstration the Old McDonald effect. The demonstration is easy to repeat. You may substitute a Coke can, or any other small can, and a wadded piece of paper for the coffee cup and ball. You may also try to hit the paper with your finger instead of a ruler, as long as you attempt to make one smooth, rapid swing parallel to the surface of the stand. If you are among the many people who are surprised to see themselves swing above the ball, you will be in a better position to appreciate the point of the demonstration, which is that *you cannot trust your intuitions about perceptual impressions and performance, even in over-learned skills such as hitting objects with your hand in natural conditions*. This is the main take-home message that I tried to emphasize in my presentation.

This message is relevant to other projects that were presented at the conference. For example, some simulators have displays that are so realistic that an observer gets an impression of actually being at the scene that is displayed, and scientists are attempting to analyze the determinants of telepresence (see Held, this volume). Held outlined a framework for analyzing determinants of the compellingness of these impressions, including time lags in visuo-motor tasks. The distinction between perceptual impressions and performance will be critical in this context if it turns out that the factors influencing perceptual compellingness are different than those determining proficiency of performance. Similarly, those who are studying stereopsis (e.g., Enright, this volume; Foley, this volume; Schor, this volume; and Stevens, this volume) might find different factors affecting impressions of depth and performance with 3D displays. Finally, efforts are being made to train pilots to see relative vertical separations better in collision-avoidance situations (Sherry Chappell,

personal communications, September 1, 1987). Scientists might also find here that factors influence perceptual impressions and performance differently.

My short-term goal is to alert applied and basic researchers about potential discrepancies between the determinants of perceptual impressions and performance that could affect their research and instrumentation designs. For now, scientists will have to watch their step on a case-by-case basis since there are no empirically founded principles that would enable general predictions. The last section of this commentary will turn to my long-term goal of providing a foundation for such predictions.

## THEORETICAL AND EMPIRICAL FOUNDATIONS FOR PREDICTING RELATIONSHIPS BETWEEN STIMULUS INFORMATION, PERCEPTUAL IMPRESSIONS, AND PERFORMANCE

This volume contains three hypotheses that propose a framework within which to investigate the many-to-one relationship that exists between stimulus information, perceptual impressions, and performance: (1) the Perception Plus Transformation hypothesis (see Foley, this volume); (2) the Dual Mode of Visual Representation hypothesis (see Bridgeman, this volume); and (3) the Ecologically Insulated Event Input Operations (EIEIO) hypothesis. Figure 1 illustrates all three. They all begin with conversion of distal information, which is in the environment, into proximal information, which is at the interface between the environment and the sensory system. According to the Perception Plus Transformation hypothesis, proximal information is converted into abstract symbolic representations that result in perceptions and sensory-guided performance. But sometimes, according to this model, the representations are transformed before they influence performance. According to the Two Modes of Visual Representation hypothesis, the proximal pattern is converted into two representations that are determined by separate neural pathways. One of these representations mediates perceptions and verbal responses, the other mediates motor responses. Finally, according to the EIEIO hypothesis, the proximal pattern is converted into multiple abstract symbolic representations. One of these is formed by general input operations that mediate perceptual impressions and some sensory-guided behaviors. The others are formed by specialized input operations, EIEIOs, which mediate specific sensory-guided skills. The general input operations are the most robust in that they are adapted to operate optimally over the entire range of variability to which the system is exposed. This robustness is gained at the expense of efficiency and accuracy in any given situation. For example, the processing efference-based and light-based information in a well-lit, structured environment might be less efficient than the processing of light-based information alone, but this strategy would protect an organism that is suddenly confronted with a situation in which the light-based information is reduced. In contrast, EIEIOs develop to serve sensory-guided skills optimally in a specific context. These input modules are extremely powerful in that context, but are very vulnerable to failures outside that context.

I originally postulated the existence of EIEIOs to account for highly skilled sensorimotor performance of athletes, pilots, and astronauts. I then realized that they might also apply to more common, highly practiced skills such as grasping, catching, or hitting objects within arm's reach. The ball and cup demonstration is consistent with this possibility. Accordingly, perceptual impressions in that situation are mediated by general input operations that are relatively robust to the elimination of binocular information because redundant monocular information is also processed. In contrast, hitting responses in that situation are mediated by an EIEIO. The results suggest that this particular input module is more dependent upon binocular information. This strategy

might have provided the EIEIO with greater efficiency in one common situation, but sacrificed robustness in other situations.

It is one thing to consider that a select few of our species, such as athletes and pilots, develop specialized event input operations to service their extremely high level sensorimotor skills. It is quite another to suggest that we all do it to control ordinary skills such as grasping, catching, and hitting in our everyday lives. An implication of the latter possibility is that the domain of perception with respect to perceptual impressions, and the domain of perception with respect to sensory-guided performance, might be more distinct than we had realized. Consequently, we might have to modify our analytic approaches to these domains. Past analyses of the nature and determinants of perceptual impressions have yielded fundamental principles such as the laws of organization. Do these principles apply to the input operations that underlie sensory-guided performance? The present considerations suggest that this question must be answered by empirical tests rather than by assumptions. The uniqueness of the EIEIO hypothesis is in the heuristic implications for such tests.

After my presentation I was asked to explain how the EIEIO hypothesis differs from other modularity models. I will conclude by answering this question. A salient feature of the EIEIO model is that it includes more than one abstract, symbolic representations of space, only one of which corresponds to perceptual impressions. As illustrated in Fig. 1, other models include that characteristic. Summary comments on this conference provided a historical context for consideration of such models (see Stark, this volume). In light of these comments and my own attempts to trace historical roots, I believe that the EIEIO model has not only a novel name, but also unique heuristic merits that will become clearer when more data are collected. In checking out the five premises that are outlined in my paper, I will be testing ideas for which there are no other tests that I have been able to find. The unprecedented experiments will focus on ways in which practice of a sensory-guided skill can reconfigure the way in which input operations utilize proximal information. Two types of experiments are suggested: one that analyzes existing skills, as was done in the ball and cup demonstration, and one that examines the learning of new sensory-guided skills. The focal questions concern the constants and variables of adaptive input operations that underlie relationships between stimulus information, perceptual impression, and performance. The processes underlying the laws of organization might be examples of processes that are universal and constant across all input operations. But, as noted earlier, the EIEIO hypothesis indicates that such possibilities must be tested rather than assumed.

Given the limited scope of this commentary, I can only paint in broad strokes the kind of tests that are suggested to me by the EIEIO heuristic. The tests that I am planning were greatly influenced by work summarized by Marr (1982). He provided detailed models of lower visual processes at three levels of explanation: (1) computation, (2) representation and algorithm, and (3) hardware (neural) implementation. In contrast, models of higher processes were limited to the computational level and were much less developed. A sharp decline in detail occurred in modeling the transition from a viewer-centered frame of reference (two and one-half-dimensional sketch) to a three-dimensional frame of reference based on the shape itself. Marr stated that an obstacle to more detailed modeling of these higher processes is the difficulty of discovering "what systems and schemes are actually used by humans...at present I see no empirical way of approaching this type of problem. It seems to be much more difficult to design experiments to answer questions at these rather high levels of analysis than at the lower ones...Designing a successful empirical approach to such questions would represent a major breakthrough." Experiments that gave major insights into lower-input operations were often based on dramatic perceptual impressions, such as those created by Julesz's random dot stereograms or by Ullman's rotating cylinder demonstrations. Higher

operations, such as those that underlie object recognition and/or localization, are much more difficult to capture with such demonstration because of the variable and complex relationships that exist between stimulus information, perceptual impressions, and performance.

In order to account for these many-to-one relationships, Marr proposed a model that bears directly upon the present considerations. He suggested that a single, two and one-half-dimensional sketch is constructed in order to serve all sensory-guided systems, and that different systems process this abstract, symbolic representation according to different rules to suit different purposes. Since the data employed in testing the nature and determinants of the two and one-half-dimensional sketch were based on perceptual impressions, Marr's model can be interpreted as a Perception Plus Transformation model.

The EIEIO is similar in suggesting distinct input modules for different purposes, but the EIEIO model does not assume common operations for all modules up to the level of a two and one-half-dimensional sketch, or up to any other abstract representation. Instead, the EIEIO model leaves open for testing the possibility that separate input models already diverge at the initial sampling of the proximal pattern, which is defined at the interface between physical information and sensory receptors before abstraction processes begin.

This contrast between models suggests a starting point for testing. My plan is to use displays similar to those that have cast light on processes that yield a two and one-half-dimensional sketch. One such display is Ullman's counterrolling cylinders, which consists of a sequential presentation of a set of frames. Each frame is a random set of dots, and that is how each frame appears when it is presented alone. The relationship between frames, however, is highly structured such that the frames present a screen containing successive orthographic projections of two concentric cylinders that are counterrotating. When the frames are presented at the appropriate rate, observers see counterrotating cylinders. This perceptual impression was Ullman's main response measure. I will modify the display in order to manipulate monocular versus binocular viewing, stereopsis, texture gradients, brightness gradients, and other information about the screen's orientation and distance. I will also add both verbal and motor response as well as more task demands and response measures, such as more detailed reports of perceptual impressions as measured by ,Epstein and Park (1986), measurements of forced-choice recognition, measurements of viewer-centered surface orientation and distance by means of alignment of an unseen body part with the surface, measurements of object-centered surface orientation by means of comparison with a standard object, and measurements of accommodation and convergence. An initial step will be to replicate the Old McDonald effect in this context and to pursue other discrepancies between perceptual impressions and performance, including recognition and visuomotor coordination. In addition, I will try to create such discrepancies by selectively manipulating sources of information during training session on different tasks.

An important phase will be testing opposing predictions of Perception Plus Transformation models and the EIEIO model. For example, control over separate sources of information will enable precise manipulations of the degree of veridicality of perceptual impressions. Perception Plus Transformation models will be supported whenever recognition or localization responses are related to perceptual impressions by a transformation rule; the EIEIO hypothesis will be supported whenever sensory-guided performance and perceptual impressions vary independently. Finally, the Two Modes of Visual Representation hypothesis will be tested by comparing verbal and motor responses.

The proposed empirical approach that was suggested by the EIEIO heuristic is a hybrid of methods traditionally used to measure perceptual impressions such as the constancies, and methods that have been used to analyze cognitive processes such as stages of processing in pattern recognition. The approach is aimed at two goals; (1) to provide a data base for inferring the systems and schemes that determine perceptual impressions and sensory-guided performance; and (2) to advance spatial instrument technology by enhancing our ability to understand, predict, and control the many-to-one correspondence that often exists between stimulus information, perceptual impressions, and performance.

## REFERENCES

Epstein, W.; and Park, K. (1986): Continuous optical transformations do not elicit unique perceptual description. Percept. Psychophys., 40, 365-469.

Marr, D.(1982): Vision, New York: Freeman.

## A. Perception Plus Transformation
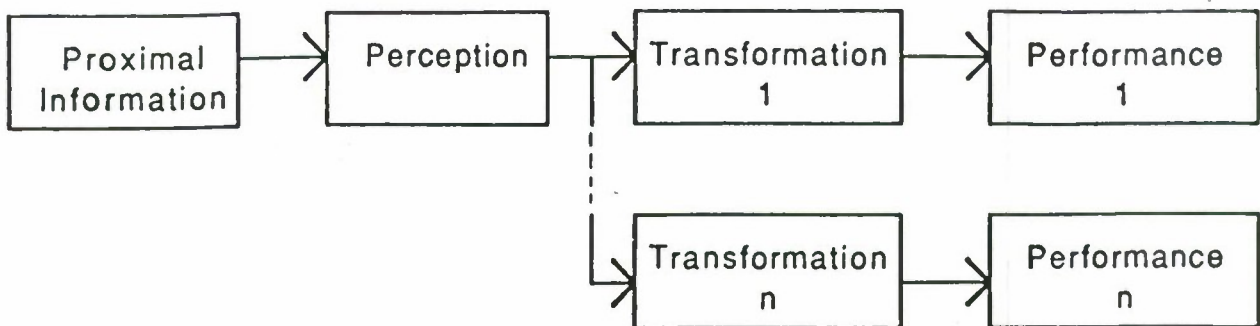
```
┌──────────────┐      ┌──────────────┐      ┌──────────────┐      ┌──────────────┐
│   Proximal   │──►   │  Perception  │──►   │Transformation│──►   │ Performance  │
│ Information  │      │              │      │      1       │      │      1       │
└──────────────┘      └──────────────┘      └──────────────┘      └──────────────┘

                                            ┌──────────────┐      ┌──────────────┐
                                      ──►   │Transformation│──►   │ Performance  │
                                            │      n       │      │      n       │
                                            └──────────────┘      └──────────────┘
```

## B. Two Modes of Visual Representation

```
┌──────────────┐      ┌──────────────┐      ┌──────────────┐
│   Proximal   │──►   │  Perception  │──►   │   Verbal     │
│ Information  │      │              │      │ Performance  │
└──────────────┘      └──────────────┘      └──────────────┘

                                            ┌──────────────┐
                                      ──►   │   Motor      │
                                            │ Performance  │
                                            └──────────────┘
```

## C. Ecologically Insulated Event Input Operations (EIEIOs)

```
┌──────────────┐   ┌──────────────┐   ┌──────────────┐   ┌──────────────┐
│   Proximal   │──►│   General    │──►│  Perception  │──►│ Behaviors in │
│ Information  │   │Input Operations│  │              │   │General Context│
└──────────────┘   └──────────────┘   └──────────────┘   └──────────────┘

                   ┌──────────────┐   ┌──────────────┐   ┌──────────────┐
                ──►│    EIEIO     │──►│    EIEIO     │──►│   Skill  1   │
                   │      1       │   │Representation 1│  │  Context 1   │
                   └──────────────┘   └──────────────┘   └──────────────┘

                   ┌──────────────┐   ┌──────────────┐   ┌──────────────┐
                ──►│    EIEIO     │──►│    EIEIO     │──►│   Skill  n   │
                   │      n       │   │Representation n│  │  Context n   │
                   └──────────────┘   └──────────────┘   └──────────────┘
```
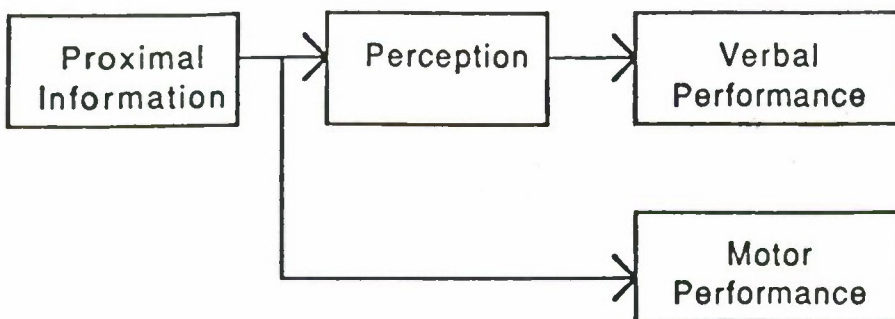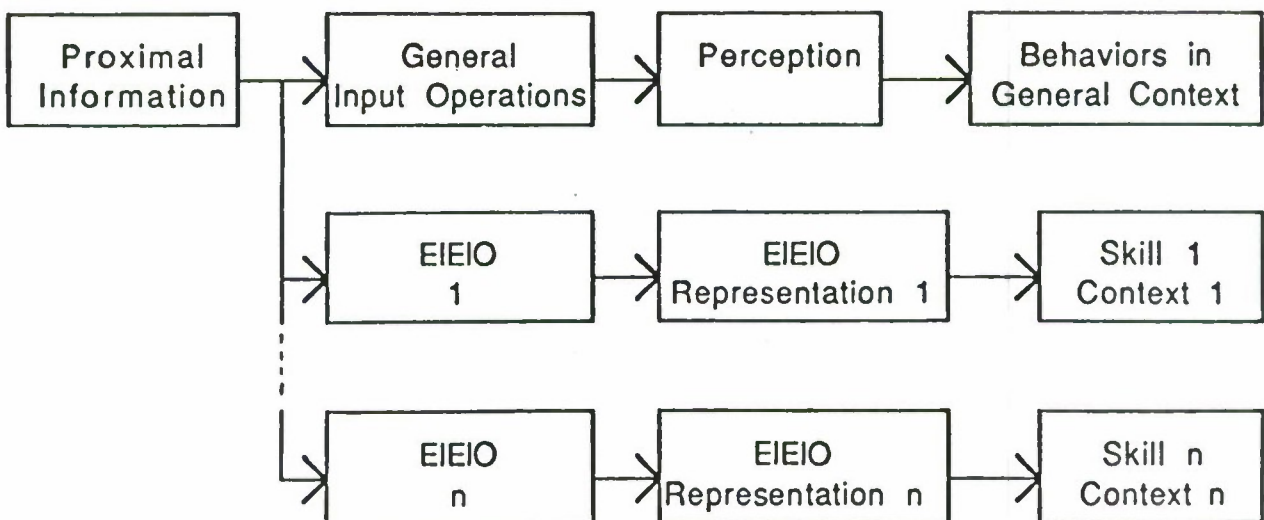
Figure 1.– Hypotheses proposing framework within which to investigate the many-to-one relationship existing between stimulus information, perceptual impressions, and performance.

# VEHICULAR CONTROL

# SPATIAL DISPLAYS AS A MEANS TO INCREASE PILOT SITUATIONAL AWARENESS

Delmar M. Fadden, Dr. Rolf Braune, and John Wiedemann
Boeing Commercial Airplane Company
Seattle, Washington

At least three elements influence the performance of an operator who must make a system achieve a desired goal: (1) the dynamics of the system itself, (2) the nature of the possible inputs, and (3) the means whereby the operator views the information concerning the desired and actual state of the system (e.g., Poulton, 1974; Wickens, 1984; and Wickens, 1987). In conventional airplanes manual control involves the coordination of "inner loop" controls. In this task the pilot is responsible for continuous manipulation of the controls to compensate for disturbances. Primary displays (fig. 1) provide the several essential flight parameters which the pilot is required to monitor, interpret, transform, and integrate.

It has long been recognized that intense concentration is necessary for a pilot to achieve high tracking performance using only "raw data." The underlying need for such concentration stems from the effort necessary to obtain timely error, error rate, and control input information in each of the three flight axes. Precision instrument approaches often have higher minimums if a suitable flight director or autopilot is not available and in use. Most pilots have come to depend on these aids. Some pilots express doubt about the precision of their own tracking ability any time they are unavailable.

Flight directors, which came into widespread airline use in the 1960s, aid the pilot in achieving improved performance by combining the error and error rate information; producing a control command appropriate to the situation. This command is then compared with the existing control input and the difference displayed as a steering command. The generation of the steering command entails automation of several logical and mathematical operations. Of course the pilot must set up the proper task for the flight director to perform and must follow the steering commands. In typical applications the automation is sufficiently complete that the pilot has no *required* intermediate data-interpretation role beyond that of recognizing and following the steering command.

While use of the flight director improves performance in precision tasks, it does not significantly reduce the continuous attention demands imposed on the pilot. Use of a path-following autopilot mode automates the process one step further by coupling the steering command to the control surfaces. Relieved of the continuous steering requirement, the pilot is able to devote more time to other tasks.

Both flight directors and autopilots achieve impressive performance gains. A side effect of these gains is a reduction in the necessity for the pilot to maintain a high level of awareness of the elements pertinent to the control task; namely the path error, error rate, and control input. To be sure, all modern aircraft present these parameters and most airline operating procedures dictate that the pilot monitor them while using either the flight director or autopilot. However, the monitoring task is fundamentally different from that of developing a control input given only "raw" data. In particular, the dynamic decision-making demands of the monitoring task are much lower than those of the control task.

Spatial displays, together with enhanced manual control, offer an opportunity to achieve the same high performance achieved with autopilots and flight directors while improving the pilot's overall situational awareness, particularly during flight tasks other than final approach. This is accomplished by revising the split of responsibilities between the pilot and the aircraft automation.

In transport operations, the need to alter the velocity (flightpath angle, track angle, or speed) is much less frequent than the need to compensate for wind effects, turbulence, configuration changes, and speed changes. In terminal area operations, the number of required velocity changes may be an order of magnitude or more lower than those attitude changes necessary to maintain a velocity. Furthermore, the needed velocity changes are typically separated by many seconds. By assigning the velocity-hold task to the basic flight-control system, the majority of the attitude adjustments can be made transparent to the pilot. This type of control frees the pilot from the continuous attention requirement of attitude steering while maintaining the pilot's direct involvement in airplane guidance.

Spatial displays make it possible for the pilot to be directly involved in developing the path error information and in selecting the specific tactic to be employed in correcting the error. To make this practical, current position and velocity information must be displayed in a consistent context. Operational displays based on work done at Boeing, NASA Langley, RAE-Weybridge, and other places have shown that a map display, with track angle and speed shown by means of predicted future positions, provides a suitable context.

The first generation of commercial airline spatial displays are in operation on the Boeing 757 and 767 and the Airbus A-310. These displays take the form of CRT maps with various types of integral predictors (fig. 2). The format consistency of these displays is quite high and pilot acceptance has been exceptionally good. The CRT maps are used for planning and assessing all types of lateral maneuvers. Direct manual aircraft control is still accomplished by reference to a separate attitude instrument, but virtually all of the decisions to maneuver laterally can be made looking at information contained in the map display.

The success of the map display and the potential for flightpath angle and track angle control to be used on the next generation of commercial aircraft encouraged us to consider expanding the role of spatial displays. Data from the NASA Aviation Safety Reporting System identifies altitude-related errors as the single largest category of reported problems (Reynard, Ames Research Center, 1987, personal communication). While the immediate causes of the reported errors are quite varied, we see a common thread emerging. The pilot's awareness of the vertical flight situation in most instances does not match the reality of the flight plan, the ATC clearance, or the equipment setup. A spatial display should be an ideal means of improving the pilot's vertical situation awareness (Baty, 1976).

For most transport flight operations except takeoff and landing, the tracking accuracy required of the pilot is at least an order of magnitude higher for the vertical task than for the lateral task. Typical tracking-performance goals as perceived by the pilot away from final approach are ±50 ft of altitude and ±0.5° of a VOR radial. At 40 n. mi. from the VOR station, ±0.5° corresponds to over ±2000 ft. At this point the accuracy ratio is 40:1. Even on final approach the vertical accuracy requirements exceed the lateral by at least 2:1. If a conformal 3-D display were used with sufficient resolution to satisfy the vertical task, the pilot would be overworked laterally. This concern, along with the difficulty of presenting future trend information in a forward-looking

display, lead us to concentrate on a separate 2-D side-view display for the majority of vertical situation information (Grunwald, 1980; and Filarsky, 1983).

Some past aircraft programs have referred to the attitude display as a vertical situation display. We prefer to use the more conventional terminology, ADI (attitude director indicator) or PFD (primary flight display) for the forward-looking display of attitude information and other fundamental flight data. We refer to a side-looking or profile display as a vertical-situation display and expect that the pilot would obtain the majority of overall vertical situation awareness from this display (fig. 3).

Over the past 2 yr we have been exploring ways of developing a useful and effective means to portray vertical-situation information. There are a number of practical problems which narrow the possible format options for vertical flight information. The remainder of this paper will outline the larger hurdles and indicate what progress has been made in solving them.

Three issues appear to be fundamental to the development of a successful vertical situation display:

1. Handling of the large difference in resolution requirements between the longitudinal and vertical flight tasks.

2. Determination of the appropriate level of control information to be contained in the instrument.

3. Selection of a display context which will be intuitive to the pilot and provide useful assistance for on- and off-path vertical maneuvering.

## SCALING ISSUES

The disparity which exists between vertical and lateral resolution requirements applies as well to vertical and longitudinal information. In fact, since time constraints are seldom tighter than a minute or more, the difference in resolution requirements can be well in excess of two orders of magnitude. With this large a difference, equal vertical and horizontal display scaling is clearly impractical. By using a flightpath predictor we have been able to achieve a balance between vertical tracking performance and the desired path preview capability.

Initial test results indicate that when the vertical situation is presented spatially, a steady increase in mean deviation from an optimal descent occurs as scale resolution is decreased (fig. 4). However, even the largest deviation is significantly less than the lowest mean without the spatial graphics. This result could be attributed to the difference in the tactics the subject pilots employed to accomplish the task under the two presentations. Without the graphics the pilots had to mentally integrate various analog quantities according to their own individual rules of thumb. As can be seen in figure 5, this results in an overall greater deviation from the optimal descent strategy and more variance among the individual pilot deviations. When given a spatial presentation of the situation, the subject pilots employed similar path-following tactics, resulting in greater tracking precision and a lower-rated workload level.

The fact of unequal scales causes the angle representations on the display to be exaggerated vertically. Through a preliminary test series we found that scale differences of as much as 20:1 do not have a negative influence on typical airline flying tasks. Obviously aircraft with significantly greater climb or descent capabilities than transports would encounter difficulty at lower-scale ratios. What appears more important to the pilots is that the longitudinal scaling of the side-view display and the map display be congruent so that the rate of movement between the two is compatible.

Another result from our initial investigations reveals that a digital readout of altitude takes on added importance as scale resolution is decreased (fig. 6). In seeking the proper balance between scale resolution for precision and scale range for preview, it was shown that a digital readout of altitude provides a good vernier indication while the graphics provides the necessary "big picture" overview. The graphic spatial information is effective in drawing the pilot's attention to the digital readout when precise control is needed.

## CONTROL ISSUES

In all of today's transport aircraft, manual control is exercised using the attitude display with follow-up reference to the situational displays. This is the case for map-display-equipped aircraft as well. Laterally the track angle is two integrations removed from aircraft roll rate, over which the pilot has direct control. The resulting time delay between control input and map response is too long for track angle to provide primary inner-loop feedback to the pilot. Even when lateral acceleration is used to create a prediction of the dynamic path which will be flown, the pilot's primary control feedback comes from the bank indication on the attitude indicator.

Vertically the conventional control parameter is pitch rate. This term is separated from flight-path angle by a single integration and some higher-order dynamics. For transports this places the flightpath response on the order of 1-2 sec behind the control input; long enough to be useless as the primary feedback term for most situations and short enough to interact negatively with pitch feedback. The primary dynamic term in the vertical situation is flightpath angle. Furthermore, flightpath angle, rather than pitch attitude, can be readily assessed in terms of the geometry or energy conditions of the vertical situation. If the response dynamics of flightpath angle were not so close to that of pitch attitude, the separation of control and situation assessment, which works very well in the lateral case, could be established for the vertical case as well.

Beginning with experimental work on the Boeing SST in the late 1960s and continuing through the early phases of the NASA TCV program, we became convinced that if flightpath angle, along with suitable situational reference information, is available to the flight crew, the crew will attempt to use it for control. Without good matching of the control and display dynamics, pilot workload may well increase.

If a flightpath-angle command-control system is in use, it is possible to display the flightpath which will be held. This term can be made as responsive as necessary to support the pilot's need for timely information. If a more conventional control system is used, a filter with appropriate lead compensation can be added to quicken the dynamics of the flightpath angle information (Bray, 1981).

The key situational element which makes control possible is the flightpath prediction based on flightpath angle. Remove the prediction and control reverts to conventional techniques. However, without the prediction, the usefulness of the display for enhancing current situational awareness is dramatically reduced. Even maintaining a constant altitude is difficult without the prediction. Thus the question about the desired level of control information is not an independent issue. If the display is to be useful, it must contain dynamic flightpath information. The presence of such information means that the display will be used for control. The real issue, then, is how to match the control and display dynamics to the information-processing capabilities of the pilot.

## DISPLAY CONTEXT

The third fundamental issue has to do with matching the frame of reference of the display to that of the pilot. The vertical component is straightforward. However, the options for the horizontal component are more complex. If information concerning the planned route of flight were always available and current, then distance along the route would be a good choice. However, the planned route is not always available. Furthermore, one of the more important uses of the display is during operations when the airplane is intentionally away from the planned path.

For these situations a narrow slice ahead of the airplane would be more useful. In either case close coordination between the vertical and horizontal situation displays is essential.

Development work aimed at clarifying the format orientation issue is now under way. We expect to have an understanding of the major tradeoffs late this year.

## CONCLUSIONS

Our experience raises a number of concerns for future spatial-display developers. While the promise of spatial displays is great, the cost of their development will be correspondingly large. The cost goes well beyond time and materials. The knowledge and skills which must be coordinated to ensure successful results is unprecedented. From the viewpoint of the designer, basic knowledge of how human beings perceive and process complex displays appears fragmented and largely unquantified. Methodologies for display development require prototyping and testing with subject pilots for even small changes. Useful characterizations of the range of differences between individual users is nonexistent or at best poorly understood. The nature, significance, and frequency of interpretation errors associated with complex integrated displays is unexplored and undocumented territory.

Graphic displays have intuitive appeal and can achieve face validity much more readily than earlier symbolic displays. The risk of misleading the pilot is correspondingly greater. Thus while we in the research community are developing the tools and techniques necessary for effective spatial-display development, we must educate potential users about the issues so they can make informed choices. The scope of the task facing all of us is great. The task is challenging and the potential for meaningful contributions at all levels is high indeed.

# REFERENCES

Baty, D. L. (1976) Rationale and Description of a Coordinated Cockpit Display for Aircraft Flight Management, NASA Technical Memorandum 3457, November 1976, Ames Research Center, Moffett Field, CA.

Bray, R. S.; and Scott, B. C. (1981) A Head-Up Display for Low Visibility Approach and Landing. Proceedings AIAA 19th Aerospace Sciences Meeting, January 12-15, St. Louis, MO.

Filarsky, S. M.; and Hover, G. W. (1983) The Command Flight Path Display. Proceedings of the Society for Information Displays, pp. 196-197.

Grunwald, A. J.; Robertson, J. B.; and Hatfield, J. J. (1980) Evaluation of a Computer Generated Perspective Tunnel Display for Flight Path Following. NASA Langley Tech Report #1736.

Poulton, E. C. (1974) Tracking Skills and Manual Control. New York: Academic Press.

Wickens, C. D. (1984) Engineering Psychology and Human Performance. Columbus, OH: Charles Merrill.

Wickens, C. D. (1987) The Effects of Control Dynamics on Performance. In K. Boff and L. Kaufman (Eds.) Handbook of Perception and Performance. Dayton, OH: University of Dayton Research Institute.
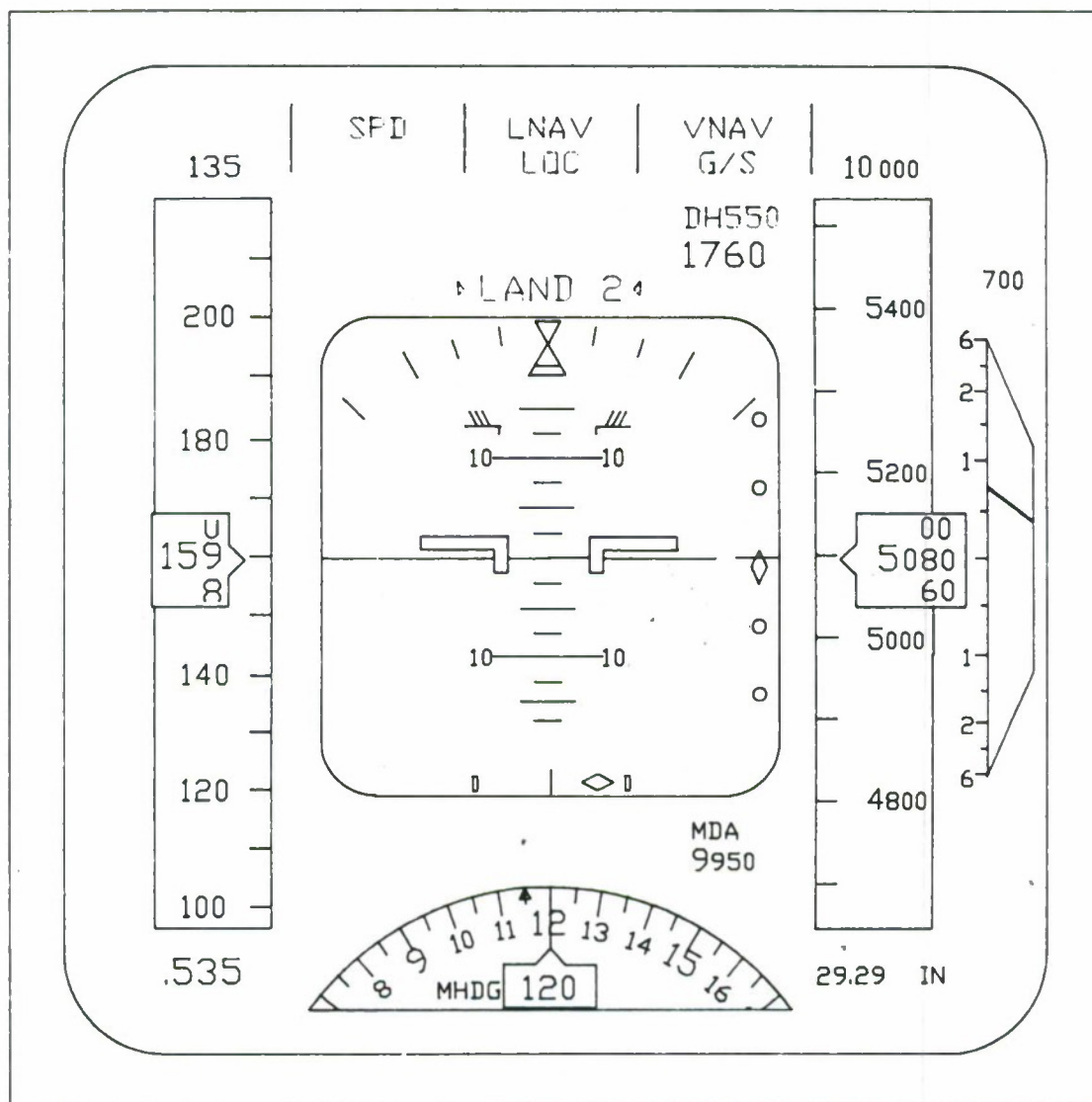
Figure 1.– Primary flight display.
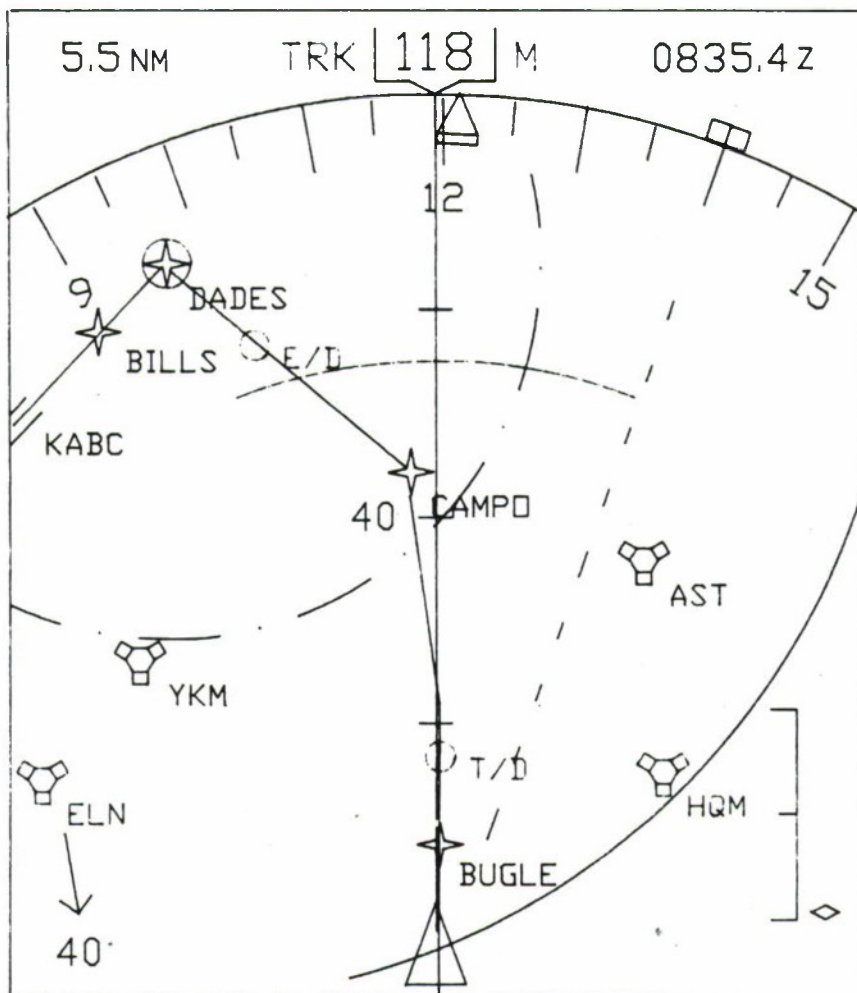
Figure 2.– Map display.
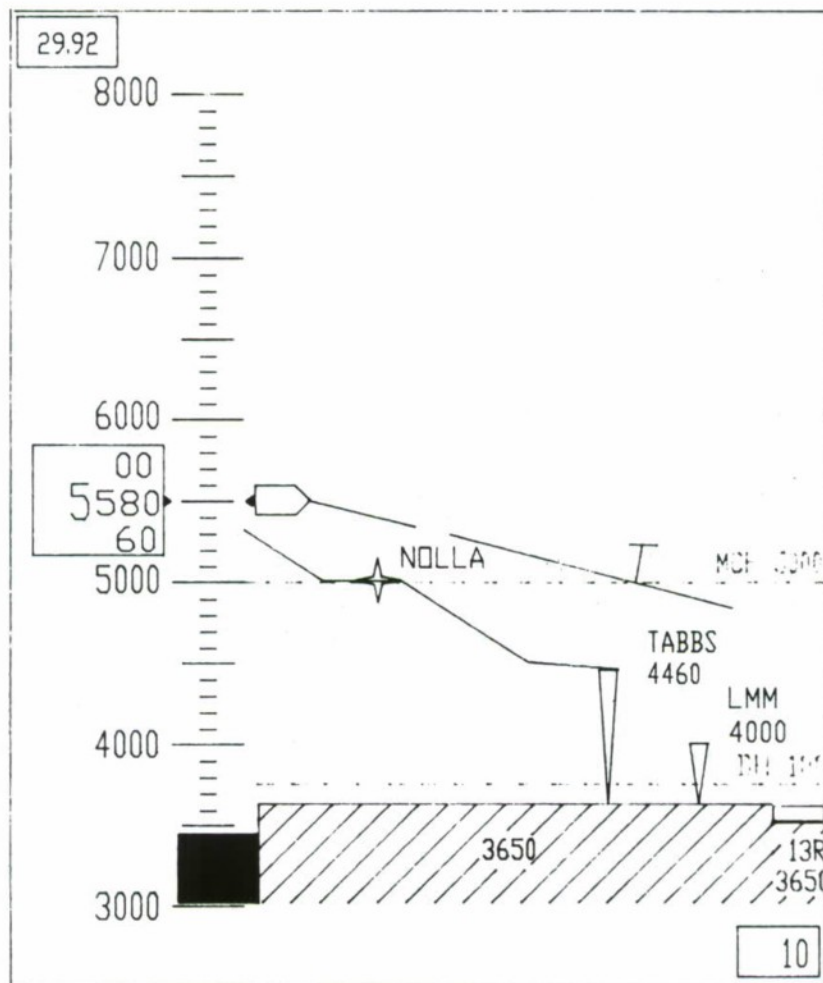
Figure 3.– Vertical situation display.
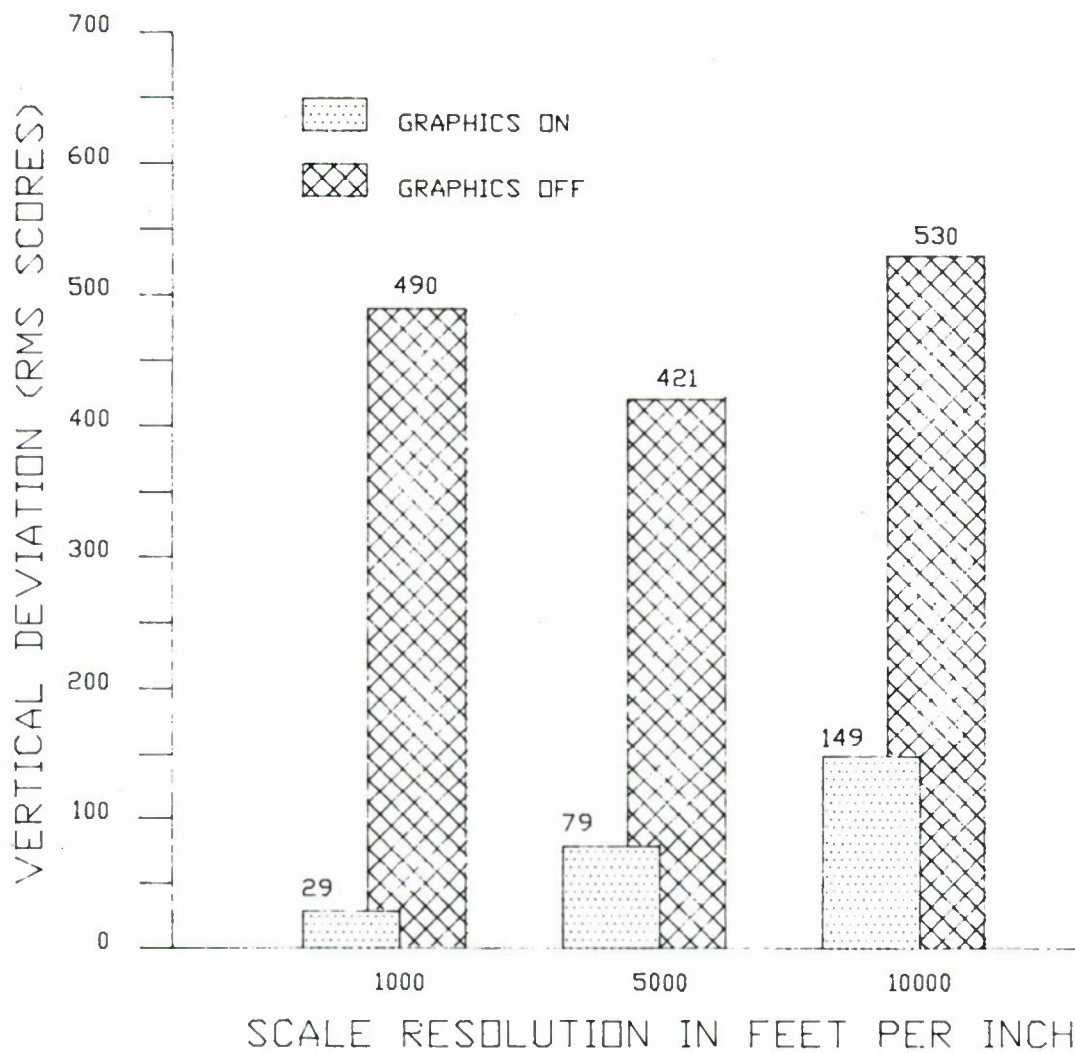
Figure 4.– Overall means for graphics × resolution.
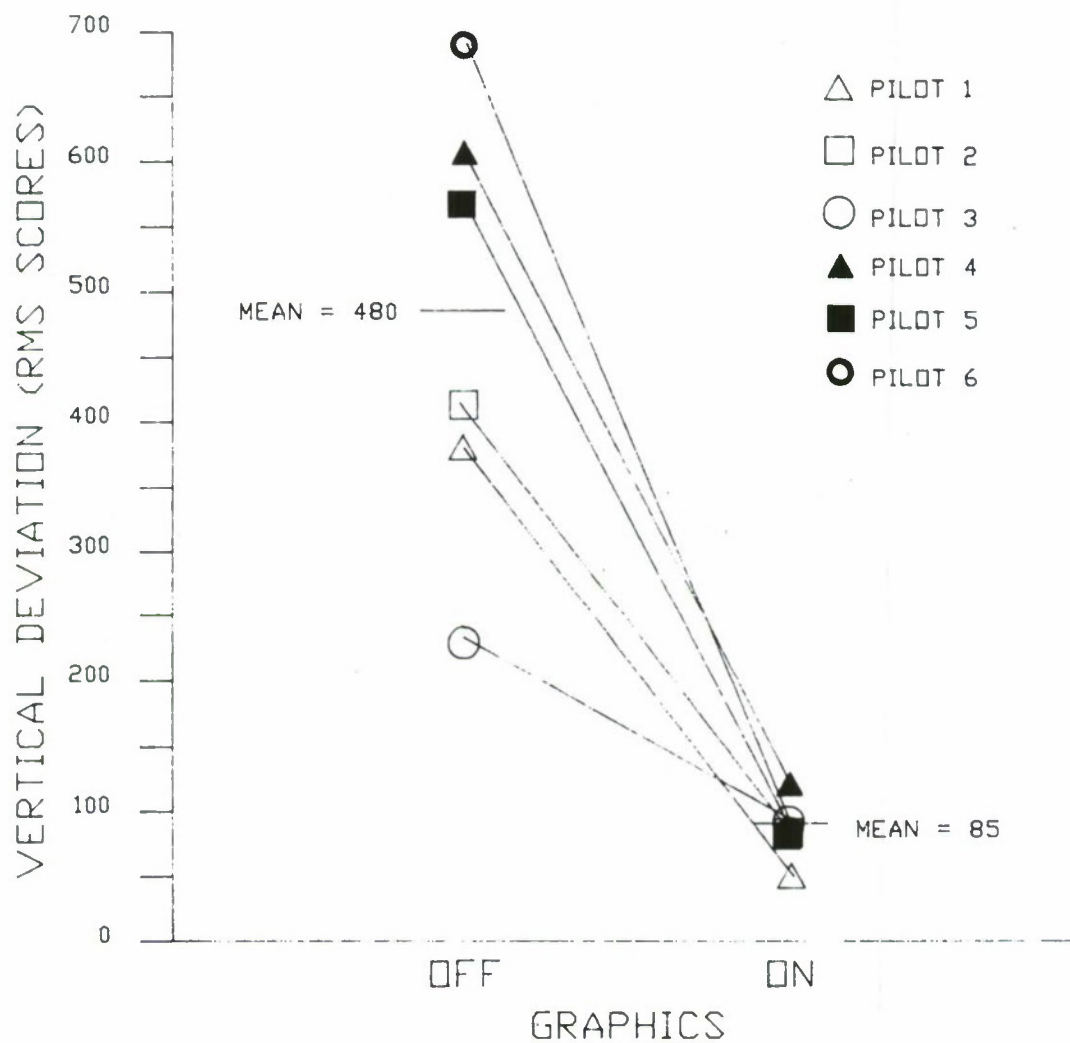
Figure 5.– Overall means for graphics × pilots.
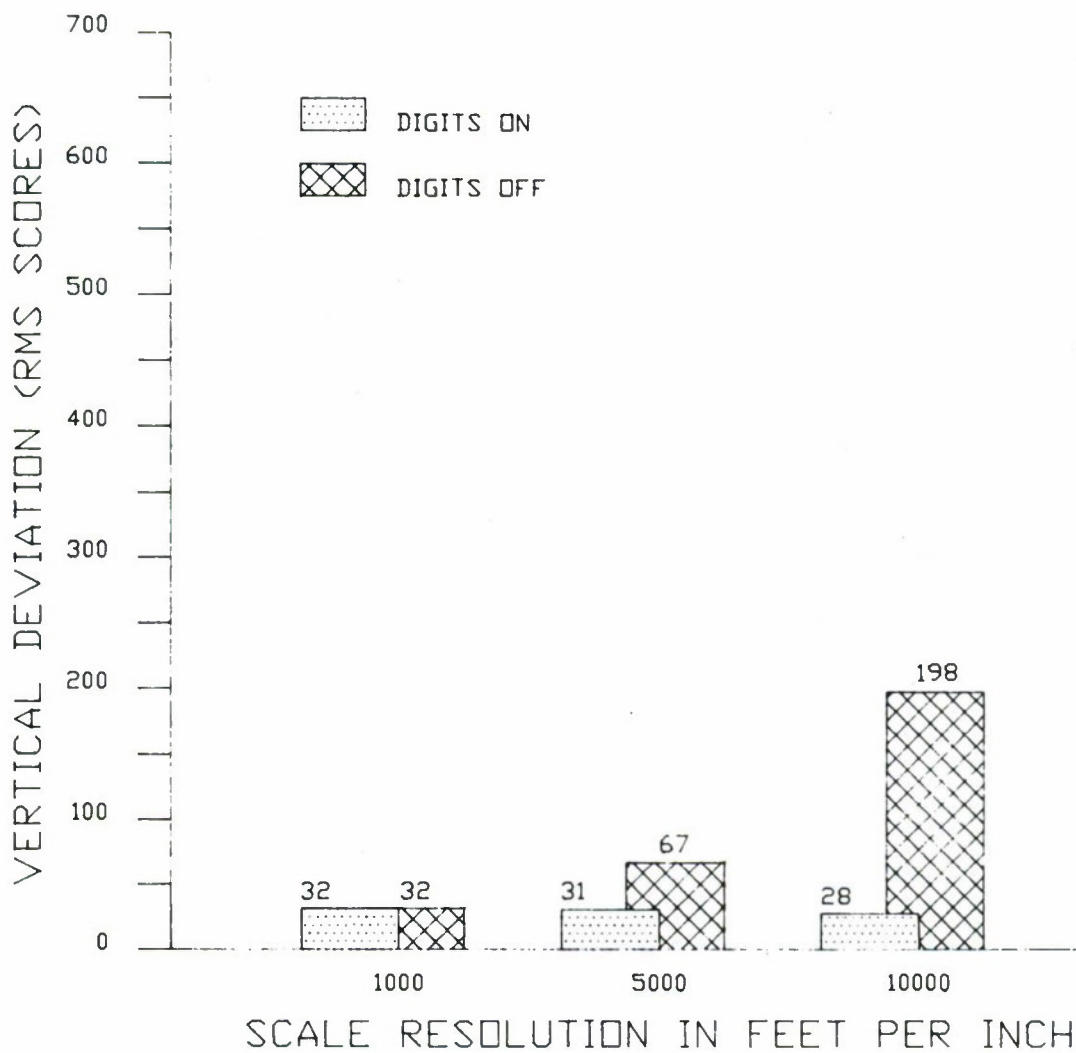
Figure 6.– Level-off segment (graphics on) means for digits × resolution.

# A COMPUTER GRAPHICS SYSTEM FOR VISUALIZING SPACECRAFT IN ORBIT

Don E. Eyles
Charles Stark Draper Laboratory
Cambridge, Massachusetts

## SUMMARY

To carry out unanticipated operations with resources already in space is part of the rationale for a permanently manned space station in Earth orbit. The astronauts aboard a space station will require an on-board, spatial display tool to assist the planning and rehearsal of upcoming operations. Such a tool can also help astronauts to monitor and control such operations as they occur, especially in cases where first-hand visibility is not possible. This paper describes a computer graphics "visualization system" designed for such an application and currently implemented as part of a ground-based simulation. The visualization system presents to the user the spatial information available in the spacecraft's computers by drawing a dynamic picture containing the planet Earth, the Sun, a star field, and up to two spacecraft. The point of view within the picture can be controlled by the user to obtain a number of specific visualization functions. The paper describes the elements of the display, the methods used to control the display's point of view, and some of the ways in which the system can be used.

## INTRODUCTION

This paper describes a computer graphics display system designed to facilitate the visualization of spacecraft operations in Earth orbit.

The system was originally developed as a component of the Space Station Simulator project at the Charles Stark Draper Laboratory. The purpose of this simulator is to assess the flying qualities of space station configurations, and to provide a software framework within which to develop control-system concepts applicable to space stations. Computer graphics were added to the simulator to provide qualitative information about the progress of the simulation, and to allow for a man-in-the-loop capability. As time went on it became evident that the displays required by engineers working on the ground might also be valuable to astronauts working aboard a space station.

To be able to carry out unanticipated tasks with resources already in Earth orbit is part of the purpose of a permanently manned space station. Operations will be required which cannot be rehearsed by the astronauts using ground-based simulators because the need for them arose after the crew was launched into space. On-board capabilities must exist to allow the crew to plan such orbital operations and to train themselves to execute them. In addition, the space station crew must perform a sort of air-traffic-control function in keeping track of other spacecraft operating nearby, and must control not only the space station itself and its movable appendages, but also free-flying spacecraft associated with the space station, including spacewalking astronauts.

The display described in this paper, if attached to suitable mission and simulation software aboard a space station, can support both the on-board simulation capability and the real-time monitoring of operations. I shall speak of the visualization system as an on-board display instrument, with the understanding that its capabilities arose from, and are also applicable to, ground-based engineering simulation purposes.

The display is called a "visualization system" because it is a system designed to aid the user in visualizing a three-dimensional situation in space. In the terminology established for this conference, the visualization system fits in somewhere between a "spatial display" and a "spatial instrument." Like a spatial display, the system presents the user with an unembellished, undistorted image of a spatial situation. Like a spatial instrument, the system requires a degree of interaction with the user, who must control the point of view from which the image is drawn. Perhaps the best description of the visualization system is as a spatial instrument which can present a variety of spatial displays under the control of the user.

The existing implementation uses display equipment that produces "wire-frame" objects whose "hidden" parts are visible. Although in some cases wire frames may remain preferable, actual use aboard a space station will require flight-qualified display hardware capable of rendering solid objects with shading and shadowing.

I shall describe the elements of the scene created by the visualization system, discuss the means by which the point of view within the scene is controlled, and finally describe some of the specific ways in which the system can be used. A more detailed description of the visualization system is available in reference 1. A short published description with color illustrations is available in reference 2.

## ELEMENTS

The principal elements of the display created by the visualization system are the planet Earth, the Sun, a field of stars, and one or two spacecraft. The planet Earth is drawn as a sphere made up of latitude and longitude grid lines and a map showing the outlines of major land masses and principal cities. Other Earth-fixed features such as circles indicating coverage from tracking sites can be added. The Earth is drawn from data expressed in a geodetic or Earth-fixed coordinate frame; that is, a frame of reference which moves with the Earth. The Sun is drawn, not to scale, as a yellow asterisk with 24 points. A star field of 123 stars is also drawn, and is valuable for two reasons. First, showing the stars in their correct astronomical positions provides a realistic star background for maneuvers being monitored or simulated, and allows maneuvers to be planned which may be dependent on the availability of specific navigational stars. Second, stars provide a motion cue when the point of view is rotating with respect to inertial space.

The visualization system also contains, in the present implementation, up to two spacecraft. One is often the space station. Each spacecraft may consist of a core and one or two movable appendages such as solar panels. For a spacecraft with thrusters, an exhaust plume is drawn when a jet is fired. Because the space station is not yet fully defined, and because other spacecraft may need to be represented, the visualization system allows spacecraft to be defined as an assemblage of simple cylindrical and plate elements. The visualization system also contains information such that a cylinder which is meant to represent an established type of module, for example a habitat

module, can be given detail to make its appearance more realistic. In the case of an unusual or unknown spacecraft, simple cylinders and rectangular plates can be used to build up an image.

Additional minor elements of the visualization system include a gnomon, always drawn in the upper left corner of the square window occupied by the display, which indicates the orientation of the local-vertical, local-horizontal (LVLH), frame of reference pertaining to the principal spacecraft. The visualization system also has the capability of drawing a buoy, a yellow three-dimensional cross, which may be used to represent present a spacecraft of unknown configuration, or to mark a spot in space, as, for example, a nominal position to be returned to after a maneuver. There are some other minor embellishments which apply to specific ways in which the visualization system can be used, and these are discussed later.

## Information Requirements

The Sun, stars, Earth, and spacecraft together form a sort of computerized orrery. The system is set into motion by computed transformations and positions which are used to locate each element in its proper relative position, either for the present time (for monitoring), or for some future time (for simulation). Besides initialization information specifying the configurations of the spacecraft that are to be drawn, the visualization system requires the following dynamic information from the simulation or mission software to which it is attached:

- Position of spacecraft center of mass.

- Position of spacecraft center of mass with respect to spacecraft structure.

- Position of spacecraft appendages.

- Attitude (orientation) of spacecraft.

- Jet firing information for spacecraft.

- Sun position.

- Transformation relating the Earth-fixed coordinate system to a reference inertial coordinate system.

- Transformation relating the spacecraft LVLH coordinate system, a frame which moves with the spacecraft and is defined in terms of its position and velocity, but not its attitude, to the reference system.

- Transformation relating the spacecraft "body" coordinate system, which is fixed with respect to the spacecraft's structure, to the LVLH frame.

New values for each quantity are required for each frame drawn by the visualization system.

For a given time, the relationships defined by this information form a scene which is representative of a real situation and not under the control of the user of the system. The point of view

within the scene, however, can be controlled by the user to accomplish various specific visualization functions.

## Control By the User

The point-of-view characteristics which are under the control of the user are the following:

- The coordinate system with respect to which the point of view will be defined.

- The origin, i.e., the object or point which is to occupy the center of the picture.

- The distance from the eye to the chosen origin.

- The line-of-sight vector, expressed in the chosen coordinate system, from the eye to the chosen origin.

- The angular field of view of the image presented.

In the present implementation all characteristics are dynamically under the control of the users as they use the display, with the exception of field of view, which is defined at initialization time. The user-controllable characteristics are input by means of an alphanumeric display and keystroke language based on the method used in the space shuttle. An analog dial and joystick may also be used in controlling point-of-view distance and direction. Although normally each characteristic is explicitly controlled, canned combinations can be provided so that certain favorite set-ups can be obtained with a minimum of keystrokes. Figure 1 shows the alphanumeric display page used to control the visualization system point of view.

The point of view coordinate system may be chosen from among the usual frames of references used in space applications. These include an inertial frame locked to the stars; an Earth-fixed frame which moves with the planet Earth; the LVLH frame which moves with the spacecraft, but is independent of its orientation; and a "body" frame which is locked to the spacecraft structure. When more than one spacecraft is included, the LVLH and body frames pertaining to each are available, although of course when the spacecraft are near each other their LVLH systems are not significantly different. All frames except the reference inertial are rotating coordinate systems. Additional coordinate systems can easily be added to the structure.

A second aspect of the point of view which is under the control of the user is the point upon which the display is centered. An early lesson in the design of the visualization system was that when the point of view is allowed to maneuver independently, it was easy to lose track of the object of interest. As a result, the point of view is normally centered on some chosen point. The choice of "origin" consists of the center of the Earth, the centers of mass, body coordinate system origin, or the crew station of either spacecraft, and the midpoint between the centers of mass of two spacecraft.

Having chosen an origin and a coordination system, the user must choose a line-of-sight vector. The line of sight is controlled by numerically specifying a unit vector expressed in terms of the chosen coordinate system, or by manipulating a joystick which is attached to the system. The line-of-sight distance, the distance between the "eye" and the chosen origin, may be controlled

numerically or by means of an analog dial. Distances between zero and 500,000 km (continuous) are permitted in the present implementation. A negative distance may be specified, but the usefulness is limited because that puts the chosen origin behind the eye.

The point of view may be thought of as looking inward from a spot on a sphere. The sphere is stationary with respect to the chosen coordinate system, it is centered on the chosen origin, and its radius is the chosen distance. The point of view's position on the sphere is specified by the line-of-sight vector.

The angular field-of-view of the display is also under the control of the user, although in the present implementation in the space station simulator, the field of view must be chosen ahead of time and is not subject to real-time modification. Fields of view between 10° and 90° are allowed. The most usual choice is 40°. While this angle does not correspond to the actual angle subtended by the display window when looked at from the usual viewing distance, it does roughly correspond to the field of view of the normal photograph taken with a medium length lens, and is satisfactory to most users.

## Ways of Using the Visualization System

The visualization system is a general system which can present the scene resulting from any combination of the available coordinate systems, origins, and lines of sight. The following are some of the specific ways in which the system can be utilized:

Chase plane views– The view that would be available from an imaginary chase plane flying alongside can be obtained by choosing the LVLH framework, an origin centered on the spacecraft of interest (or midway between two spacecraft of interest), and a line of sight and distance such as to achieve the desired view, whether from ahead, the side, behind, above, or below. Such a point of view can be useful when visualizing docking and berthing operations in which two spacecraft come together or separate. It can also be useful simply by presenting an "out of spacecraft" view of a single spacecraft, such that the spacecraft's location relative to the Earth in the background, its orientation, and the position of its movable appendages are simultaneously apparent. Figure 1 shows a chase plane view in which an OMV approaches a satellite to pick it up.

Pilot's-eye views– By selecting the body coordinate system of a given spacecraft, setting origin to "crew station," and choosing a distance of zero, the point of view can be placed in the driver's seat of any spacecraft, even an unmanned one for which an imaginary crew position is defined. Such views can serve a number of purposes, such as assessing what will be seen from the crew station window during a planned maneuver (including star availability and the problem of solar glare), presenting views that are not available in real life because there is no suitable window, and providing an on-board perspective for unmanned spacecraft which may be remotely controlled from the space station. If coupled to suitable simulation software, this point of view also allows the rehearsal of operations to be conducted by an astronaut using a Manned Maneuvering Unit, such as satellite capture. Figure 2 presents the view from a point behind the space station hatch to which the shuttle will dock. Such a view represents a "synthetic window" providing visibility in a case where spacecraft structure may preclude an actual window. (An illustration of the fact that the visualization system includes special cases equivalent to existing instruments is the fact that a pilot's eye view looking forward, perhaps with the horizon in view, corresponds to the stylized pattern presented by the attitudes reference instrument known as the 8-ball or artificial horizon.)

Whole-Earth views– The visualization system permits point-of-view distances large enough that the entire Earth is visible. Because at such distances the spacecraft appear as points of light, a capability called "rescale" is available which vastly expands each spacecraft (and shrinks the Earth), to produce a not-to-scale cartoon view in which both the position and orientation of the spacecraft are apparent. Such a point of view is useful for following a rendezvous operation in which the spacecraft may start out on opposite sides of the planet. For example, a whole-Earth view looking along the Y axis in LVLH coordinates shows the view normal to the orbital plane. Z-axis views in the Earth-fixed or inertial framework show the Earth from its polar axis. In the inertial case the Earth will be seen to rotate during 24 hr. Figure 3 shows a scene in which two spacecraft are viewed from a polar axis point of view.

Isolating a factor– Another way of using the visualization system allows the effect on a spacecraft of some single factor to be isolated. Such a capability might come into play when a new control system is to be tested on-board before being given control of the space station. The spacecraft is drawn twice at the same location and time. One image represents the spacecraft as it actually appears in real time, the other represents a simulated version of the same spacecraft as if it were controlled by the new control system under test. Divergences between the two images will illustrate performance differences attributable to the new system.

Roam capability– In most cases it is desirable to center the point of view on the object of greatest interest. The "roam" capability can be selected to remove that constraint and allow the point of view to maneuver independently within the framework established by selected coordinate system and origin. During a roam, the point-of-view orientation is controlled by a joystick and a dial can be used to creep forward or backward. When the joystick is deflected a reticle is drawn at the center of the screen to facilitate pointing at the object of interest. The reticle disappears several seconds after the joystick is released to afford an unobstructed view. The roam capability can be used to mimic a spacewalk, or EVA, by roaming within the spacecraft body coordinate frame. (However, control is geometric, and the orbital dynamics of an EVA are not simulated in this case.) The roam capability may be most important for inspecting the spacecraft's structure (as known to the computers) but, for example, the view from Boston or Los Angeles could be obtained by letting the point of view roam within the Earth-fixed frame.

The visualization system is not limited to the capabilities described. It can present any view that can be specified using the point-of-view variables under the control of the user. This can include points of view that are probably nonsensical. An example would be an Earth-centered view in the spacecraft body coordinate system. If the spacecraft is spun, the planet appears to gyrate in such a way that the spacecraft is kept in the same orientation.

## CONCLUSION

The central strategies employed in designing the visualization system were, first, to use a picture to make available to the user the extensive information available in the space station's computer system; and second, rather than design a number of special-purpose instruments, to create a general display from which specific capabilities can be obtained by controlling the point of view in various ways.

It may be useful, in conclusion, to contrast the visualization system to concepts such as the "virtual cockpit" designed to assist the pilots of high-performance aircraft. While the virtual cockpit enhances the pilot's perceptual effectiveness, the "visualization system" enhances the crew's operational effectiveness.

The distinction follows from the dissimilar missions. The mission for which the virtual cockpit is designed may last only the few seconds it takes for a jet aircraft to carry out an attack. The pilot's success and survival depend on efficiency during this period. The virtual cockpit takes a single point of view and enhances its perceptions by introducing labels, speed posts, threat indicators the terrain itself, and so forth. The attack pilots might appreciate a view of themselves as seen by the target, but the exigencies of the combat situation require instead that they stay within themselves.

The visualization system is also designed to enhance the pilot's effectiveness, but in this case the mission may last months and, despite the high absolute velocities, the relative speeds are often closer to sailboats than to jets. On the other hand, space is a place with no up or down, or rather a variety of ups and downs, depending on the particular situation. The visualization system responds by providing a tool that is suitable for the on-board planning and rehearsing that will be part of a long mission, and which offers a way of visualizing operations as they appear in several shifting frames of reference.

Aboard a space station, the pilot is sitting at a console which may face in an arbitrary direction and may be without a window. Split-second reactions are seldom necessary. There are no weather problems. What is necessary is the ability to plan and then to monitor spatial operations which may be hard to see and hard to visualize. For this case, the ability to assume a God's-eye view and follow the orbits leading to rendezvous, to fly alongside in a phantom chase plane, to take the vantage point of an imaginary window in your own spacecraft, or the viewpoint of another, perhaps unmanned satellite, may prove to be useful.

## REFERENCES

1. Eyles, Don : A computer graphics system for visualizing spacecraft in orbit, Charles Stark Draper Laboratory, R-2069, May 1988.

2. Eyles, Don: Space station thrillers unfold at Draper lab. Aerospace Amer., October 1986, 38.
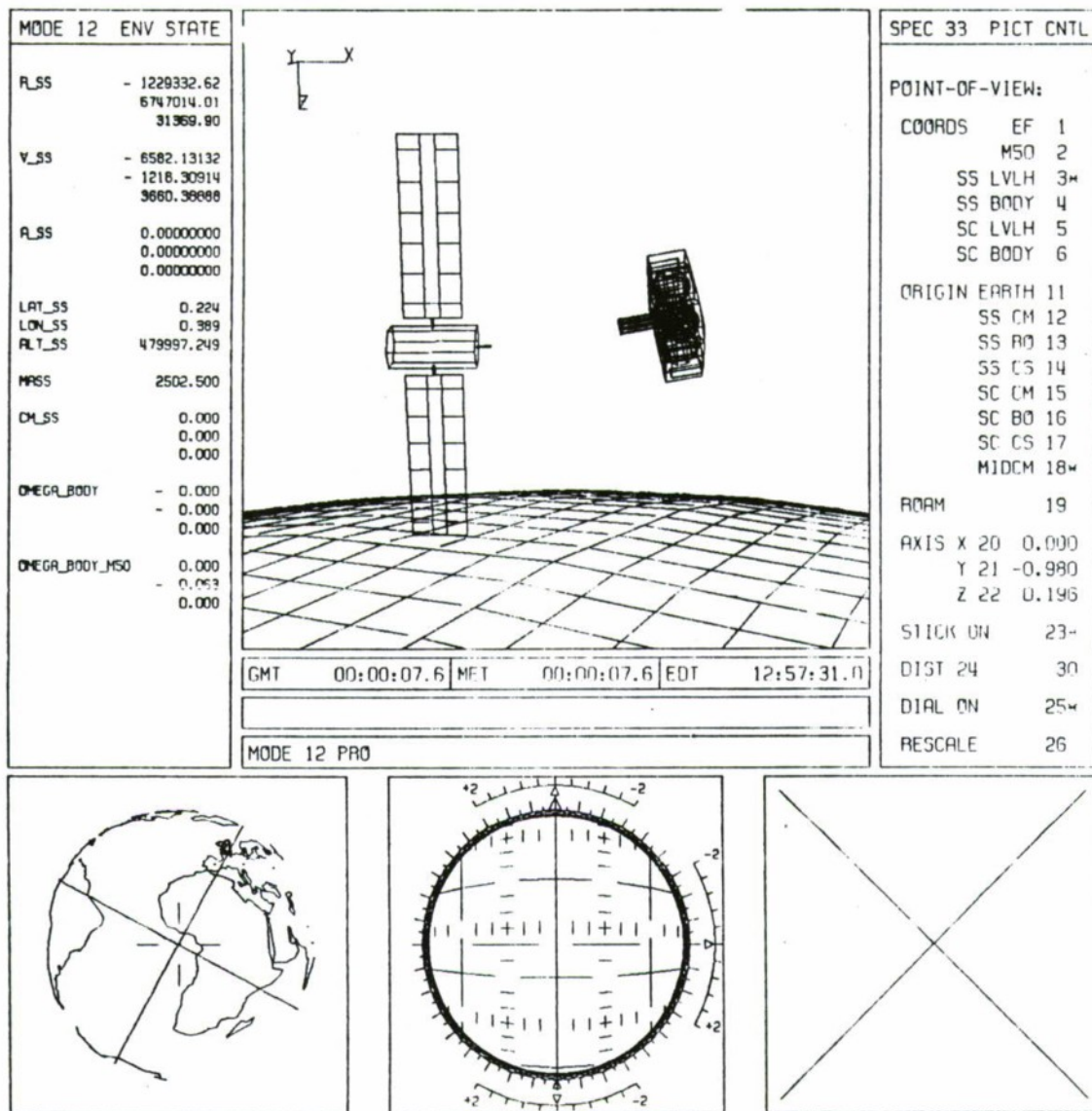
Figure 1.– The complete space station simulator display. The visualization system forms the square window at upper center. The alphanumeric display page used to control the point of view is at upper right, simulation data are displayed at upper left, and special-purpose displays such as an orbit position indicator (OPI) and an attitude director indicator (ADI) are below. The visualization system shows an orbital maneuvering vehicle (OMV) nearing a satellite which it wishes to grapple, as it would be seen from an imaginary "chase-plane" flying beside them.
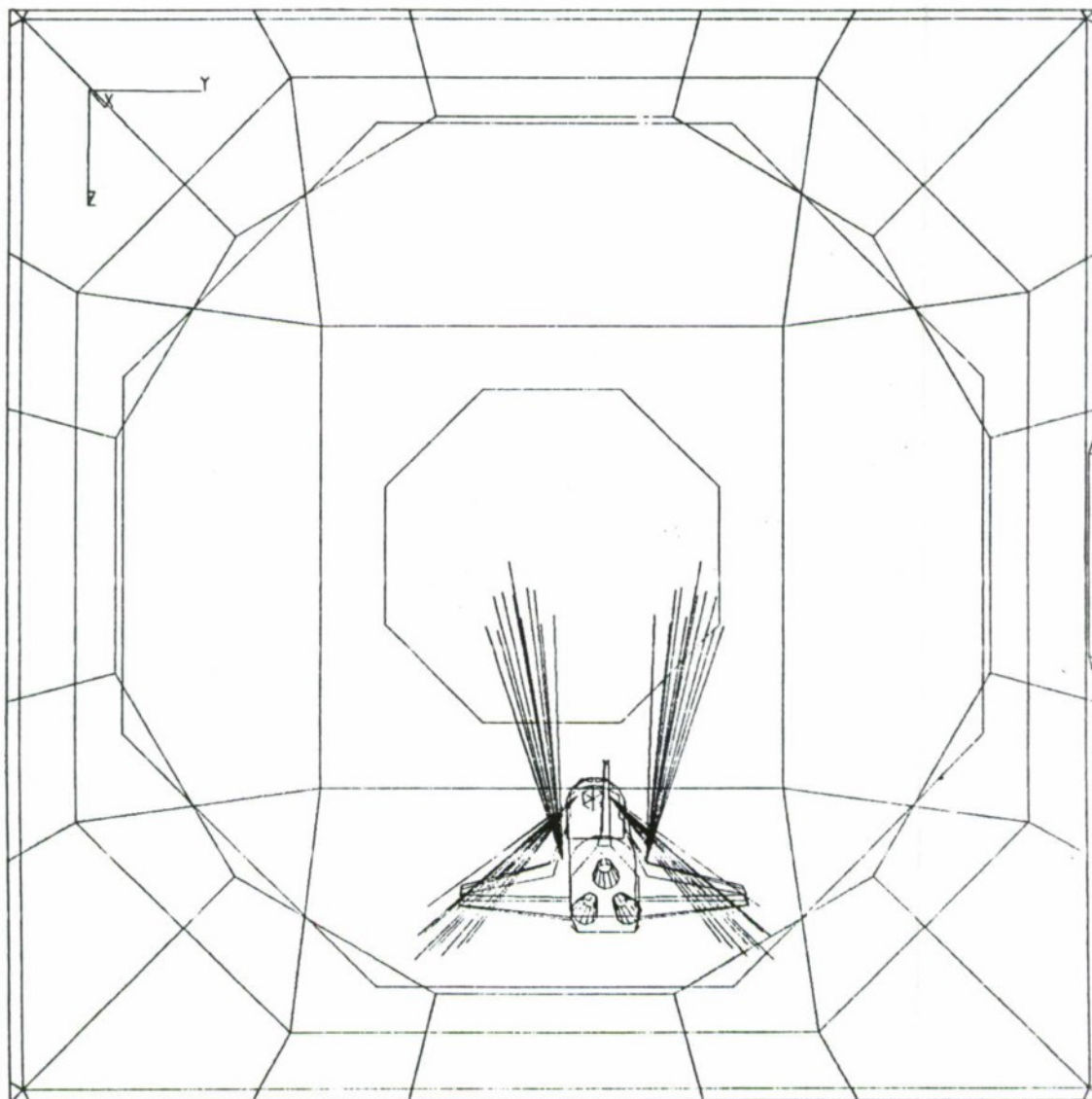
Figure 2.– In this view the point of view has been locked to the body coordinate system of the space station and located just inside the shuttle docking hatch, looking in a forward direction. At a distance of approximately 150 m a space shuttle fires maneuvering jets to reach an attitude for docking. Such a point of view can be used to assess window visibility for upcoming operations, or, as in this case, to provide a synthetic window where none exists.
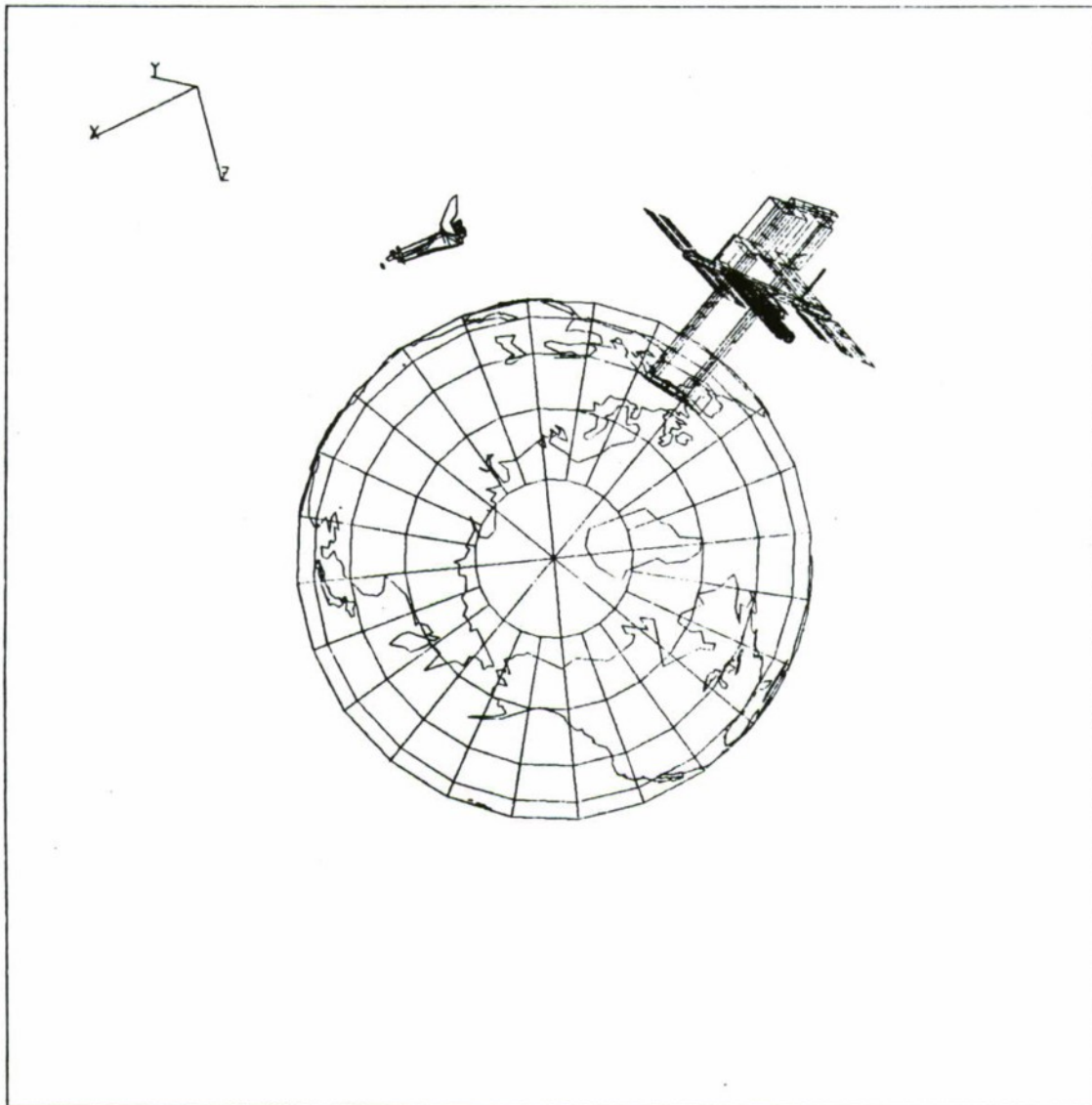
Figure 3.– In this view the point of view has been located 30,000 mi from the center of the Earth and directly above the north pole. Two spacecraft are shown, the dual-keel space station and a space shuttle. The RESCALE option has been selected and therefore the spacecraft sizes are exaggerated. Such a point of view allows the positions and attitudes of multiple spacecraft to be simultaneously visualized, as might be desirable during a rendezvous maneuver.

# INTERACTIVE ORBITAL PROXIMITY OPERATIONS PLANNING SYSTEM

Arthur J. Grunwald and Stephen R. Ellis
NASA Ames Research Center, Moffett Field, California

## ABSTRACT

An interactive, graphical proximity operations planning system has been developed which allows on-site design of efficient, complex, multiburn maneuvers in the dynamic multispacecraft environment about the space station. Maneuvering takes place in, as well as out of, the orbital plane. The difficulty in planning such missions results from the unusual and counterintuitive character of relative orbital motion trajectories and complex operational constraints, which are both time-varying and highly dependent on the mission scenario. This difficulty is greatly overcome by visualizing the relative trajectories and the relevant constraints in an easily interpretable, graphical format, which provides the operator with immediate feedback on design actions. The display shows a perspective bird's-eye view of the space station and co-orbiting spacecraft on the background of the station's orbital plane. The operator has control over two modes of operation: (1) a viewing system mode, which enables him or her to "explore" the spatial situation about the space station and thus choose and frame in on areas of interest; and (2) a trajectory design mode, which allows the interactive "editing" of a series of way-points and maneuvering burns to obtain a trajectory which complies with all operational constraints. Through a graphical interactive process, the operator will continue to modify the trajectory design until all operational constraints are met. The effectiveness of this display format in complex trajectory design is presently being evaluated in an ongoing experimental program.

## INTRODUCTION

The future space station environment will include a variety of spacecraft co-orbiting with the space station in close vicinity. Mostly, these spacecraft will be "parked" in a stable location with respect to space station, i.e., they will be on the same circular orbit. However, some missions will require repositioning or transfers to and from these spacecraft. In these cases complex types of maneuvers are anticipated which involve a variety of spacecraft which are not necessarily located at stable locations and thus have relative motion between each other.

The multivehicle environment poses new requirements which do not exist in conventional missions scenarios. The conventional scenarios involve proximity operations between only two vehicles. In these two-spacecraft missions, the scenario is in most cases optimized and precomputed in advance, and executed at the time of the actual mission. However, since the set of possible scenarios in a multivehicle environment is virtually unlimited, the future space station environment will create scenarios which might not have been precomputed and will have to be planned and executed on site. This will require an on-site planning tool which allows, through a fast interactive process, the creation of a fuel-efficient maneuver which meets all constraints set by safety rules.

The difficulties encountered in planning and carrying out orbital maneuvers originate from several causes. The first is the counterintuitive character of orbital motions as experienced in a relative reference frame. The orbital motions are expressed in a coordinate frame attached to the space station and represent relative rather than absolute motions. It would be intuitively assumed that a thrust in "forward" direction, i.e., in the direction of the orbital velocity vector, would result in a straight-forward motion. However, after several minutes, orbital mechanics forces will dominate the motion pattern and move the spacecraft "upwards," i.e., to a higher orbit. This will result in a backwards relative motion, since objects in a higher orbit move slower. Thus, a forward thrust has an effect opposite from that intended.

A second cause of the difficulty is the different and unconventional way in which orbital maneuvering control forces are applied. In atmospheric flight, control forces are applied continuously to correct for randomly appearing atmospheric disturbances, or to compensate for atmospheric drag. In contrast, spaceflight in the absence of atmospheric disturbances has a near-deterministic character. Therefore, spaceflight is mainly "unpowered" along a section of an orbit with certain characteristics. By applying relatively short impulse-type maneuvering forces at a given way-point, the characteristics of the orbit will be altered. After application of the maneuvering force, the spacecraft will coast along on the revised orbit until the next way-point is reached.

Third, multivehicle orbital missions are subject to stringent safety constraints, such as clearance from existing structures, allowable approach velocities, angles of departure and arrival, and maneuvering burn restrictions due to plume impingement. Design of a fuel-efficient trajectory which satisfies these constraints is a nontrivial task.

It is clear that visualization of the relative trajectories and control forces in an easily interpretable graphical format will greatly improve the feel for orbital motions and control forces and will provide direct feedback of the operator's control actions. Furthermore, visualization of the constraints in a symbolic graphical format will enable an interactive graphical trajectory design in which, in each iteration step, the design is modified until all constraints are satisfied.

## DESCRIPTION OF THE TECHNIQUE

### Purpose of Orbital Planning System

The purpose of the interactive orbital planning system is to enable the operator to design an efficient, complex, multiburn maneuver, subject to the stringent safety constraints of the future dense space station traffic environment, which enables a chaser to rendezvous with a target spacecraft in a given timespan. The constraints include clearances from structures, relative velocities between spacecraft, angles of departure and arrival, approach velocity, and plume impingement. Because of the complexity and counterintuitiveness of orbital motion, and the demands to satisfy strict safety rules and constraints, fuel-efficient trajectory design will be a complex and difficult task. The basic idea underlying the system is to present the maneuver, as well as the relevant constraints, in an easily interpretable graphical format. This format provides operators with immediate feedback on the results of design actions, and enables them to closely interact with the system. In an iterative process, operators will keep changing the design until all constraints are met. The methods for enabling interactive trajectory design and visualization of constraints are discussed in detail hereafter.

## Illustrative Example of a Three-Burn Maneuver

An illustrative example of a three-burn maneuver is shown schematically in figure 1, showing the situation in the orbital plane. Trajectory design can be greatly simplified by expressing the positions and velocities of co-orbiting spacecraft relative to a space-station-based coordinate system. This system $x^\circ y^\circ z^\circ$ has its origin at the the center of mass of the station and is oriented with the $x^\circ o y^\circ$ plane locally level with the surface of the Earth, with the $x^\circ$-axis in the direction of the station's orbital velocity vector and the $z^\circ$-axis pointing towards the center of the Earth. Thus, the $x^\circ o z^\circ$ plane constitutes the orbital plane. The section of the circular orbit s, followed by the center-of-mass of the space station is called the "V-bar," and the radial line r, moving outwards from the Earth center through the space station, is called the "R-bar." For the near environment of the space station, the V-bar can be considered to be straight and to coincide with the $x^\circ$-axis, and the R-bar with the $z^\circ$-axis.

The trajectory originates from relative position A at time $t = t_0$ and is composed of two way-points B and C, which specify the location in space station coordinates at which the chaser spacecraft will pass at a given time. At a way-point the orbital maneuvering system or other reaction control system can be activated, creating a thrust vector of given magnitude for a given duration, in a given direction in the orbital plane or out of the orbital plane. The duration of the burn is considered very short in comparison with the total duration of the mission. In the orbital dynamics computations this means that a maneuvering burn can be considered as a velocity impulse which alters the direction and magnitude of the instantaneous orbital velocity vector of the spacecraft.

Since the initial location A is not necessarily a stationary point, the magnitude and direction of the relative velocity of the chaser at point A is determined by the parameters of its orbit. If no maneuvering burn would be initiated at $t = t_0$, the chaser would continue to follow the relative trajectory 1, subject to the parameters of its original orbit (see dotted line in fig. 1). However, a maneuvering burn at $t = t_0$ will alter the original orbit such that the chaser will follow the relative trajectory 2, subject to the parameters of a new orbit.

In figure 1 $\underline{v}_1$ and $\underline{v}_2$ indicate the relative velocity vector of the chaser just before and after the maneuvering burn, respectively, where $\underline{v}_1$ and $\underline{v}_2$ are tangential to the relative trajectories 1 and 2, respectively. The vector difference between $\underline{v}_1$ and $\underline{v}_2$, $\underline{v}_a$, is the velocity change initiated by the burn, and corresponds with the direction and magnitude or duration at which the orbital maneuvering system is activated. Likewise, at way-point B the burn $\underline{v}_b$ alters the orbit to orbit 3.

Location C is the terminal way-point and is, in this case, the location where the target will arrive at $t = t_f$. Since the target has an orbit of its own, orbit 4, it will have a terminal velocity at $t = t_f$. The relative velocity between target and chaser is the vector difference between $\underline{v}_3$ and $\underline{v}_4$, $\underline{v}_c$. This vector determines the retroburn that is needed at the target location, in order to bring the relative velocity between chaser and target to the minimum required for the docking operation.

## Inverse Method of Solving Orbital Motion

Interactive trajectory design demands that the operator is given free control over the positioning of way-points. However, the input variables of the commonly used equations of orbital motion, as given in reference 1 and derived from references 2-4, are the magnitude and direction of the burn at $t = t_0$, rather than the position of way-points. Therefore an "inverse method" is

required to compute the values of a burn necessary to arrive at a given way-point positioned by the operator. This method is outlined hereafter.

The equations in reference 1 show how the orbital parameters of a co-orbiting spacecraft can be computed from its momentary position and velocities, relative to the space station. Thus, for a given initial relative position A with $\underline{x}(t_o)$, and an initial relative velocity $\underline{v}(t_o)$, at time $t = t_o$, the relative position and velocities of a way-point at time $t = t_1$ can be computed. However, a maneuvering burn at $t = t_o$ will cause a change in the direction and magnitude of the relative velocity vector $\underline{v}(t_o)$. As a result, the position of the way-point at time $t_1$, $\underline{x}(t_1)$ will change as well.

Consider $v_a$ and $\alpha_a$ to be the magnitude and direction of the velocity change due to the maneuvering burn. Then the relative position and velocity at $t = t_1$, $\underline{x}(t_1)$, will be a complex, nonlinear function of $v_a$ and $\alpha_a$. Consider now that the operator is given direct control over $v_a$ and $\alpha_a$ by slaving these variables directly to the x and y motions of an input device such as a control stick or mouse. An input in either x or y direction will result in a complex nonlinear motion pattern of $\underline{x}(t_1)$. Furthermore, this motion pattern will change with the initial conditions. This arrangement is highly undesirable in an interactive trajectory design process in which the operator must have direct and unconstrained control over the positioning of way-points.

It is therefore essential to give the operator direct control over the position of way-points rather than over the magnitude and direction of the burn. The inverse method by which this is accomplished computes the magnitude and direction of the burn required to bring the spacecraft from initial location $\underline{x}(t_o)$ to the way-point $\underline{x}(t_1)$ at $t = t_1$.

A Newton-Raphson method has been employed to solve this inverse problem. The operator commands the position of a way-point by means of the x-y motions of the input device. The algorithm starts with an initial guess of $v_a$ and $\alpha_a$. These values yield a computed way-point which is usually different from the commanded one. At each program update the values of $v_a$ and $\alpha_a$ are adjusted to bring the computed way-point closer to the commanded one. On the average about three to four iterations are required to bring the difference between the computed and commanded way-point effectively to zero. As the operator moves the commanded way-point around in the orbital plane, the algorithm "tracks" the commanded way-point by continuously making appropriate adjustments in $v_a$ and $\alpha_a$. As a result of this continuous adjustment, the deviation between commanded and computed way-point will remain relatively small and the Newton-Raphson scheme will operate close to the optimum. The advantage of the Newton-Raphson scheme is that convergence with this second-order technique is the best in the near vicinity of the optimum. Since the program update rate is about 15 Hz, convergence is very fast and the computed way-point is virtually indistinguishable from the commanded one.

## The Active Way-Point Concept

Although a trajectory may be composed of several way-points, only one way-point at a time, the active way-point, is controlled by the operator. The active way-point should be clearly distinguishable from the other inactive points, by conspicuous marking, highlighting, or blinking. While the position and time of arrival of the active way-point can be varied, the position and time of arrival of all other way-points remains unchanged. However, variations in the active way-point

will cause changes in the trajectory sections and way-point maneuvering burns just preceding and just following the active way-point. The on-line solution of the inverse algorithm enables these changes to be visualized almost instantaneously and provides the operator with on-line feedback on the design actions.

Although impingement constraints and approach velocity limits exist for all way-points, it is useful to limit the computation and display of these constraints to the active way-point only. This arrangement simplifies and speeds up system update computations and minimizes the symbology shown on the display. The justification for this is that the operator's attention is mainly allocated to the active way-point and its near vicinity. In a subsequent design iteration, the operator may shift the activation to a different way-point and again verify whether all constraints are met.

Since impingement constraints and approach velocity limits mainly relate to the target craft, it is useful to visualize the position of the target on the target trajectory, corresponding to the time of arrival at the active way-point. Like the active way-point itself, this position should be clearly distinguishable from other points as well.

## Way-Point Editing

The trajectory design process involves changes in existing way-points, addition of new points, or deletion of existing undesired points. An illustrative example of this way-point editing process is shown in figure 2. In the program the way-points are managed by a way-point stack, which includes an up-to-date sequential list of the position $x$, the time of arrival $t$, and the relative velocity $v$ just after initiating the burn, of all way-points.

Figure 2a shows two way-points, the initial point $x_0$ and the terminal point $x_1$. The initial way-point is defined by the initial conditions of the situation and cannot be activated or changed by the operator. The terminal way-point $x_1$ is thus the the active way-point which can be changed. The corresponding way-point stack is shown on the right. The active way-point box is drawn in bold. The relative velocity stack shows only the velocity $v_0$, which is the required relative velocity just after the burn at way-point 0, computed by the inverse algorithm, to reach point $x_1$ at time $t_1$.

Figure 2b shows the addition of a new way-point. This point is added half-way on the trajectory section just preceding the active way-point. Thus its time of arrival is chosen to be $t = 0.5(t_i + t_{i-1})$, where $i$ in this case is 1 and relates to the stack before modification. The new position, $x_1$ and relative velocity, $v_1$ are computed by the "forward" equations given in reference 1, by computing the orbital position at the new time $t$, using the existing orbital parameters previously computed with $x_0$, $v_0$, and $t_0$. The newly computed way-point position, time and relative velocity are inserted between points 0 and 1 of the stack before modification and the new way-point is chosen to be the active one. The dotted lines in figure 2 indicate variables which are transferred without modification and the encircled variables are the newly computed ones. It is important to note that since the relative velocities $v_0$ and $v_1$ are matched to the required way-points $x_1$ and $x_2$, respectively, the inverse algorithm does not need to make any adjustments.

Figure 2c shows the results of changes in the newly created way-point on the way-point stack. Since $x_1$ and $t_1$ are varied, the relative velocity at way-point 0, $v_0$ will be readjusted by the inverse algorithm and likewise the relative velocity $v_1$.

Figure 2d shows the creation of an additional new way-point. Since the active way-point prior to the addition was point 1, the new point is added half-way between point 0 and 1 and its position and relative velocity are computed with the forward method. The new values are inserted between points 0 and 1 of the stack before modification and the new way-point is again set to be the active one.

In figure 2e way-point 2 is activated. Apart from the shift in active way-point, the stack remains unchanged. The dotted line shows the the direct-path section between point 1 and point 3 without the intermediate burn at point 2. Deletion of way-point 2 will remove this point from the stack, and after that close the gap (fig. 2f). However $v_1$ has to be readjusted to fit the new direct-path section. Starting from the old incorrect value of $v_1$, the adjustment is made iteratively and on-line by the inverse algorithm.

## Operational Constraints

The multispacecraft environment will require strict safety rules regarding the clearance from existing structures. Thus, spatial "envelopes" can be defined through which the spacecraft is not allowed to pass. These spatial constraints can be visualized on the display. The operator must be able to make a clear judgment whether the planned trajectory clears the spatial constraint, or, he or she must be able to decide whether to avoid the constraint through an in-plane or an out-of-plane maneuver. However, the operator is not always able to make these judgments on the basis of one perspective aerial view or one perspective projection. In this research a graphical enhancement is used in which the spatial constraint is unambiguously presented on a time-axis display format. This format and its advantages are discussed later.

Restrictions on angles of departure and arrival may originate from structural constraints at the departure gate, or the orientation of the docking gate or grapple device at the target craft. Limits for the allowable angles of departure or arrival can be visualized on the display. In addition, the terminal approach velocity at the target might be limited by the characteristics of the grapple mechanism or the docking procedure. Limits for the allowable terminal approach velocity can be visualized as well.

Way-point maneuvering burns are subject to plume impingement constraints. Hot exhaust gases of the orbital maneuvering systems may damage the reflecting surfaces of sensitive optical equipment such as telescopes, infrared sensors, or solar panels, or may cause an undesired transfer of momentum. Maneuvering burns towards these pieces of equipment are restricted in direction and magnitude. Limits for the allowable direction and magnitude are a function of the distance to the equipment and plume characteristics. These limits can be visualized on the display.

Flight safety requires that the relative velocity between spacecraft is subject to approach velocity limits. In conventional docking procedures this limit was proportional to the range (refs. 5-7). A commonly used rule of thumb is to limit the relative approach velocity to 0.1% of the range per second. This conventional rule is quite conservative and originates from visual procedures in which large safety margins are taken into account to correct for human or system errors. Although the future traffic environment will be more complex, and will therefore demand larger safety margins, more advanced and reliable measurement and control systems will somewhat relax these demands. The effect of these developments on the allowable approach velocity limits is at present difficult to predict and so is the margin for human error to be taken into account.

In this study, the relative approach velocity is defined as the component of the relative approach velocity vector between the two spacecraft along their mutual line of sight. The limit on this relative approach velocity is a function of the range between the spacecraft. This function will depend on the environment, the task, and the reliability of measurement and control equipment, and cannot be determined at this stage. In this study a simple proportional relation has been chosen. The approach velocity limit is visualized on the display as a circle indicating the minimum range between the two spacecraft allowed for the present approach velocity. If the target craft appears within this circle, the approach velocity limit has been violated.

## DESCRIPTION OF THE DISPLAY

### Graphics System and Layout of the Display Area

The system has been implemented on a Silicon Graphics IRIS 2400 Turbo Graphics Workstation with 24 bitplanes of display memory and with a 19-inch, full-color display monitor with a display resolution of 1024 by 767 pixels. The program is named "NAVIE," which is the Hebrew word for prophet, after the prophet Elijah, who was characterized by providing trustworthy future information. Operator interaction with the system is through a two-axis, three-button mouse.

The layout of the display area is shown in figure 3. The display area has been divided into four viewports. The main area 1 is 750 by 750 pixels and areas 2,3, and 4 are 230 by 230 pixels each. Viewports 1, 3, and 4 provide information about the spatial situation about the space station, trajectories, constraints, and orbital maneuvering fuel use; and viewport 2 includes an eight-button function control panel.

### Description of Program Control Modes

The program operates in two modes. The first one, the viewing system mode, relates to the main display, which shows a perspective view of the space station and its surroundings on the background of the station's orbital plane. In the viewing system mode, the operator is able to "explore" the spatial situation about the space station and thus choose a viewpoint location and viewing direction which focuses and "frames in" on the momentary area of interest. The second mode is the trajectory design mode, in which way-points are selected, moved, added, and deleted in order to obtain a multiburn trajectory which complies with the given set of constraints.

### Viewing System Mode

The geometry of the viewing situation is shown in figure 4. The space-station-based coordinate system is $x^o y^o z^o$ with the $x^o$-axis coinciding with the orbital velocity vector, and $x^o o z^o$ is the orbital plane. Figure 4 shows the orientation of the viewing system $x^e y^e z^e$ relative to the space station system. The viewing system has its origin at point A, the $x^e$-axis coincides with the viewing direction and the image plane is perpendicular to the $x^e$-axis with the screen axes $y^s$ and $z^s$ parallel to $y^e$ and $z^e$. Point B indicates the intersection of the viewing axis with the orbital plane. Although the viewing system position, point A, and the angular orientation are defined by three

displacements and three angles, which can be all controlled independently, it is useful to constrain the motion to the following three types.

**"Tethered" motion**- In the first type of motion, the viewing system tethers about point B, which is kept fixed on the orbital grid, while the distance d between points A and B, which is the viewing range to point B, is kept constant. The tethered motion is controlled by the angles $\psi$ and $\theta$. The viewing axis $x^e$ and the axis $y^e$ are located at all times in the plane P which passes through the point B and rotates about the line CC', which is parallel to the $x^\circ$-axis, the V-bar. The line BE is also located in the plane P and perpendicular to the line CC'. $\psi$ is the angle between the $y^\circ$-axis and the line BE, and $\theta$ is the the angle between BE and the $x^e$-axis. Thus, the angles $\psi$ and $\theta$ control the obliquity of viewing along the orbital plane in the $z^\circ$ and $x^\circ$ direction, respectively. This tethered type of motion is very useful for the following reasons. (1) While the area of interest remains in the center of the display, it allows one to "explore" other possible areas of interest by changing the angles $\psi$ and $\theta$. (2) The line CC' will appear on the screen at all times as a horizontal line through the center of the display and represents a line parallel to the V-bar. Thus, while the viewing direction may change, the direction of the V-bar is at all times recognizable as the horizontal line, passing through the center of the display.

**Translational motion**- The second type of motion relates to the position of point B in the orbital plane. Here the $x^\circ z^\circ$ coordinates of point B are varied, while $\psi$, $\theta$, and d are kept constant. This translational type of motion enables the operator to move areas of interest to the center of the display.

**Ranging motion**- In the third type of motion, all parameters are kept constant except for the range d. This ranging type of motion is useful after areas of interest are located and brought into the center of the display. "Ranging-in" on the area of interest allows this area to be studied in more detail.

In the viewing system mode the operator has one-button control over the three types of motion and can "toggle" in a closed sequence from tethered motion to translational motion to ranging motion and back to tethered motion. The one-button control is useful since viewing system operations are naturally performed in a sequence of three steps, where in the first step areas of interest are searched for, in the second step the area localized during the search is moved to the center of the display, and in the third step the area is ranged in on to obtain the required level of detail.

## Trajectory Design Mode

In the trajectory design mode, the operator has control over the selection, positioning, time of arrival, addition, and deletion of the way-points which determine the trajectory. Two submodes exist: the in-plane design mode and the out-of-plane design mode. In the in-plane mode the mouse controls the $x^\circ z^\circ$ position of way-points, while the out-of-plane position $y^\circ$ remains unchanged, whereas in the out-of-plane mode the opposite is the case.

The design process starts with an initial configuration of way-points. Usually there are initially two way-points, as in the way-point editing example. The terminal point $x_1$ is the active way-point. Time of arrival at this active way-point is set to an initial value within the allowable time span of the mission. The operator has the option to increase or decrease the time of arrival at any

active way-point. The time of arrival at the terminal way-point is limited to the time span of the mission, and the one of an intermediate way-point by the time span set by the neighboring points.

As outlined previously, a convention is chosen in which a new way-point is added halfway on the time scale, on the trajectory section preceding the active way-point. The newly added way-point becomes the active one and can be moved to any desired location and its time of arrival can be set to any value within the time span determined by the neighboring way-points. However, in some cases, it is useful to "slide" the new way-point along the trajectory section connecting its neighboring way-points. The position on this trajectory section is then determined by its time of arrival only. In this mode the "locked-on-trajectory" mode, the time of arrival is slaved to the y-motions of the mouse.

The locked-on-trajectory mode is particularly useful for checking whether operational constraints between the spacecraft and the target, or other nonstationary spacecraft, are being violated. As the operator slides the way-point along the trajectory, the corresponding target position slides along the target trace as well; conflicting situations, such as a too close flyby, will be recognized immediately.

## Geometrical Enhancements; the "Time-Axis" Format

The purpose of these enhancements is to resolve ambiguities in the spatial situation by processing the spatial information and presenting it in a different format. One such format is the time-axis display which provides unambiguous qualitative and quantitative information about the out-of-plane situation and the spatial constraints.

The basic idea of the time-axis format is demonstrated in figures 5a-c. From the perspective view of figure 5a alone, it cannot be clearly determined whether the spatial constraint is violated or how the trajectory should be planned to avoid it. The view along the z°-axis in figure 5b is even less clear, because of the curved character of the trajectory. In the time-axis format of figure 5c, the out-of-plane deviation is plotted as a function of the traveled time along the path. The spatial constraints are visualized as follows. At each point on the traveled time axis, at the corresponding location on the trajectory, a line is placed perpendicular to the orbital plane. Sections of this line which are within these constraints are identified and plotted on the time-axis display of figure 5c as a set of vertical bars. Where the trajectory curve passes through these bars, the spatial constraints have been violated. Reshaping of the in-plane trajectory will alter the size and location of the constraint bars on the time-axis display. From the display it can be clearly determined whether the constraint should be avoided through an in-plane or an out-of-plane maneuver.

The format of the time-axis display used in the program is shown in figure 6. The time-axis is marked in quarters of an orbit. The shaded areas represent the nighttime section of the orbit. Both the target and the chaser trajectories are shown. It should be noted however, that although the chaser and target share the same time axis, they relate to different spatial trajectories. Therefore, the spatial constraint bars relate to the chaser trajectory only.

# Symbolic Enhancements

**Visualization of departure constraints-** Procedures at the departure gate might constrain the relative angle of departure and the magnitude of the departure burn. The in-plane constraints at the departure gate are illustrated in figure 7. The size of the burn vector is made proportional to the burn magnitude, with a scale factor of 500-m length per 1-m/sec burn, on an orbital grid with lines spaced 200 m apart. The departure constraints are satisfied if the burn vector is within the solid "bracketed" arc. This arc is specified by the arc center angle $\gamma_0$, the arc aperture $\gamma$, and the arc radius $\epsilon$. Note that maneuvering burns are expressed in terms of a velocity change rather than of a thrust force. The actual duration and thrust force of the burn depends on the spacecraft mass and the thruster characteristics.

In order to keep the display free from unnecessary symbology, it is useful to present the constraint only when it is close to being violated. If the burn vector is within the area enclosed by the dotted line in figure 7, the constraint is not drawn. The radius of the dotted arc is 80% of $\epsilon$, and the aperture angle is 10° smaller than $\gamma$.

It should be noted that the situation in figure 7 relates to a stationary departure gate. The spacecraft trajectory in this case is aligned with the burn vector. For a departure gate which moves with respect to the space station system, this will not be the case. In this case the burn vector will signify the relative direction of departure with respect to the moving gate, rather than with respect to the space station. But this vector is subject to the departure constraints and not the velocity vector of the trajectory, which is relative to the space station. Therefore, the symbology is valid for departure from a stationary as well as a nonstationary gate.

The out-of-plane constraint at the departure gate is illustrated in figure 6. The initial out-of-plane component of the burn vector has to be within the impingement constraint brackets. The out-of-plane burn scale factor is 500-m length per 1-m/sec burn. If the burn magnitude is less than 80% of the allowed maximum value, the constraint is not drawn.

**Visualization of arrival constraints-** The arrival procedures constrain the angle and magnitude of the terminal velocity vector relative to the arrival gate. The in-plane constraints at the arrival gate are visualized in figure 8. The scale factor for the relative terminal velocity vector is 500-m length per 1-m/sec terminal velocity. The arrival constraints are satisfied if this vector is within the solid arrival arc. This arc is specified by the arc center angle $\delta_0$, the arc aperture $\delta$, and the arc radius $\eta$. The arrival arc is visualized at all times.

The out-of-plane limits on the terminal approach velocity are depicted in figure 6. The approach velocity has to be within the constraint brackets. If the velocity is less than 80% of the allowed maximum value, the constraint is not drawn.

**Visualization of plume impingement constraints-** Plume impingement constraints limit the magnitude and direction of maneuvering burns. The in-plane impingement constraints of a burn given at a way-point towards the target are illustrated in figure 9. The burn-vector symbol, whose size is proportional to the magnitude of the burn, is not allowed to cross the bracketed impingement constraint arc with aperture $\beta$ and radius $\sigma$. The variables $\beta$ and $\sigma$ are a function of the distance between way-point and target $|\Delta \underline{X}| = |\underline{X}_T \sim \underline{X}|$, whose function depends on the characteristics of plume and target. In this example, B is chosen to be constant and $\sigma$ propor-

tional to $|\Delta \underline{x}|$. If the burn vector does not cross the dotted bracketed arc, the constraint is not drawn. The radius of the dotted arc is again 80% of $\sigma$ and the aperture angle is 10° larger than $\beta$.

Visualization of the approach velocity constraint- The method of visualizing the relative approach velocity limit is shown in figure 10. The relative approach velocity of the chaser towards the target is given by the vector $\Delta \underline{v} = \underline{v} - \underline{v}_T$. The line-of-sight vector of the chaser towards the target is $\Delta \underline{x} = \underline{x}_T - \underline{x}$. The relative approach velocity vector $\underline{v}_r$ is the projection of $\Delta \underline{v}$ on $\Delta \underline{x}$ and is given by

$$\underline{v}_r = (\Delta \underline{v}^T \Delta \underline{x})\Delta \underline{x}/|\Delta \underline{x}|^2 \tag{1}$$

where T denotes the transpose, or inner product. The limit on $|\underline{v}_r|$ is a function of the distance between chaser and target $|\Delta \underline{x}|$. In this example, a simple proportional relationship has been chosen. Thus, for a given approach velocity $|\underline{v}_r|$, the allowable range $\rho$ can be computed and visualized by a circle centered about the chaser's position. The approach velocity constraint is violated when the target is located within this circle. The circle is visualized when $\rho$ is greater than 80% of $|\Delta \underline{x}|$.

Orbital fuel use- The orbital fuel use is displayed in viewport 4. The orbital fuel is expressed in total m/sec velocity change rather than kg fuel mass. The actually spent fuel mass depends on the spacecraft and the thruster characteristics and will be proportional to the total velocity change. A fuel dial is shown which indicates the percentage of fuel remaining from the total amount allowed for the mission. The remaining fuel is indicated by a yellow sector, and fuel use in excess of the allowed amount is indicated by this sector turning red. In addition to the fuel dial, the percentage of fuel left and total fuel use are displayed numerically.

Trajectory time markers- Along the chaser and the target trajectories, time markers are placed at regular intervals. The time marker is a small bar, perpendicular to the trajectory, provided with a number which indicates the time in minutes after starting the maneuver. Special care is given to the automatic repositioning of the numericals after a viewing system change. The numericals are placed such that they do not "clutter" the trajectory and clearly point to the corresponding time marker.

## Computational Enhancements

Computation of the relative trajectories is a time-consuming process, which, if done at each program update, will result in an unacceptable low update rate, jerky motions, and poor control over the positioning of a way-point. This can be prevented by disabling the trajectory computations and starting them only after the operator has completed the positioning of a way-point. At each program update interval, the x and y output values of the mouse are compared with the values from the previous step. If no change has taken place, a timer is initiated. The trajectory computations are started 0.3 sec after initiating the timer. After the trajectory is computed, the computed values are stored and displayed and no further computations will take place until the next change in way-point position. The 0.3 sec delay is essential for assuring that the operator has completed the positioning process. Often, small corrections are made after the way-point has been moved the first time. Experience has shown that, in most cases, no more changes are made after a 0.3 sec delay. Sometimes subsequent changes are made after the operator has reviewed the position. These

changes are seldom made earlier than 0.5 sec after the last change and this is after the trajectory has been recomputed.

It should be noted that although the trajectory computations are subject to delay, this is not the case with the computation of variables which relate to the way-points themselves, such as maneuvering burn vectors, relative velocity vectors, and operational constraints. The computation of these variables is less time-consuming and is done at each program update interval. Continuous update of these variables is essential in order to give the operator immediate feedback of the effect of a certain design action on maneuvering burns or approach velocities.

## DISCUSSION

The proposed interactive orbital planning system should be seen as a preliminary step in determining the display format which will be useful in the dense space station environment. The examples shown here deal with the most general situation, which involves departures from, and arrival at, nonstationary locations. However, most of the co-orbiting spacecraft are likely to be "parked" on the V-bar, and thus at stationary positions. Missions with spacecraft at nonstationary positions and substantial out-of-plane motion thus represent a worst-case situation, and are chosen here to demonstrate the capabilities of interactive graphical trajectory design, rather than representing the common type of maneuver to be executed at the station.

Likewise, it is hard to predict whether the constraints used here will be relevant and realistic in the future space station environment. They predict in a broad sense the type of restrictions which are expected in the multivehicle environment, e.g., limitations on approach rates, plume impingement, and clearance from structures. It is also likely that the future environment will pose different constraints, which might originate from the specific character of a mission, like a specific scenario in which a telescope or manufacturing platform is approached and serviced.

A further restriction of the display relates to the way the orbital maneuvering system is activated. Only pure impulse maneuvering burns are considered, in which the duration of the burn is negligible with respect to the duration of the mission and in which these burns cause major changes in the relative trajectories. Station-keeping or fly-by missions, however, require a more sustained type of activation, such as periodic small burns with intervals of several seconds over a time span of several minutes. A more distributed way of activating the orbital maneuvering system can be introduced in which the operator has control over the frequency and time span of the activation. Ways should be found which enable this type of control to be activated and visualized.

A last restriction relates to the way the spatial trajectory is visualized. The perspective main view shows the projection of the actual trajectory on the orbital plane, rather than the trajectory itself. The reason for this is two-fold. The orbital trajectory, with its typical cycloidal shape, when shown without lines projected on the orbital reference plane is ambiguous and might seem to bend out of the orbital plane. This illusion results from the viewer's familiarity with objects such as a coil spring and has first been reported in reference 8. Therefore, the trajectory cannot be shown without its projection on the orbital plane. Second, the symbolic enhancements and burn vectors relate to the in-plane motion and match with the trajectory projection on the orbital plane. Thus, both the trajectory and its projection should actually be visualized. However, in a perspective plan view, i.e., viewed along the $y^\circ$-axis, both the trajectory and its projection on the orbital plane will

show up as separate curves which might be highly confusing. Therefore a compromise has been sought, in which the projection is shown together with "pedestals" placed at the way-points orthogonal to the orbital plane, which mark the actual trajectory at the way-points.

In spite of these restrictions, the proposed display clearly demonstrates the usefulness of interactive graphical trajectory design. The use of the graphical, symbolical, and computational enhancements indicates the direction in which a solution for a multivehicle environment display should be sought. A still-unanswered question relates to the degree of automatization which should be introduced in the display. Parts of the mission could be performed through the use of optimization techniques, e.g., to find the fuel-optimal way-point which clears a spatial constraint in part of the mission, or to find a way-point which satisfies the terminal constraints. However, since the solution space of a complex situation is virtually infinite, it is yet doubtful whether this mission can be performed entirely automatically. It is therefore expected that frequently occurring routine operations, such as searching the local solution space for the optimal location of a way-point, might be handed over to an optimization scheme. These solutions can be reviewed by the operator, and manually changed if necessary.

In a presently ongoing experimental program, operators are carrying out a series of design missions which vary in complexity and constraints. In a tutorial session, the operators are first familiarized with the orbital motions, orbital control methods, operational constraints, and the system control functions of the viewing system motions and way-point editing process. Each operator action is time-marked and recorded. Statistics of the viewing system actions will show "preferred" viewing situations for each condition. Review of the trajectory design actions might identify the existence of heuristic design rules which might be utilized in automated design schemes.

# REFERENCES

1. A.J. Grunwald and S.R. Ellis, "Interactive Orbital Proximity Operations Planning System," NASA TP-2839, 1988.

2. L.G. Taff, <u>Celestial Mechanics, A Computational Guide for the Practitioner</u>, John Wiley & Sons, 1985.

3. J. Kovalevsky, <u>Introduction to Celestial Mechanics</u>, Translated by Express Translation Service, Springer-Verlag, New York, 1984.

4. W.T. Thomson, <u>Introduction to Space Dynamics</u>, Dover, New York, 1986.

5. NASA Lyndon B. Johnson Space Center, "Flight Procedures Handbook," JSC-10566, Nov. 1982.

6. A.R. Brody, "Spacecraft Flight Simulation: A Human Factors Investigation into the Man-Machine Interface between an Astronaut and a Spacecraft Performing Docking Maneuvers and other Proximity Operations," M.Sc. Thesis, Massachusetts Institute of Technology, April 1987.

7. NASA Lyndon B. Johnson Space Center, "Rendezvous/Proximity Operations Workbook," RNDZ 2102, 1983.

8. S.R. Ellis and A.J. Grunwald, "A New Illusion of Projected Three-Dimensional Space," NASA TM 100006, 1987.
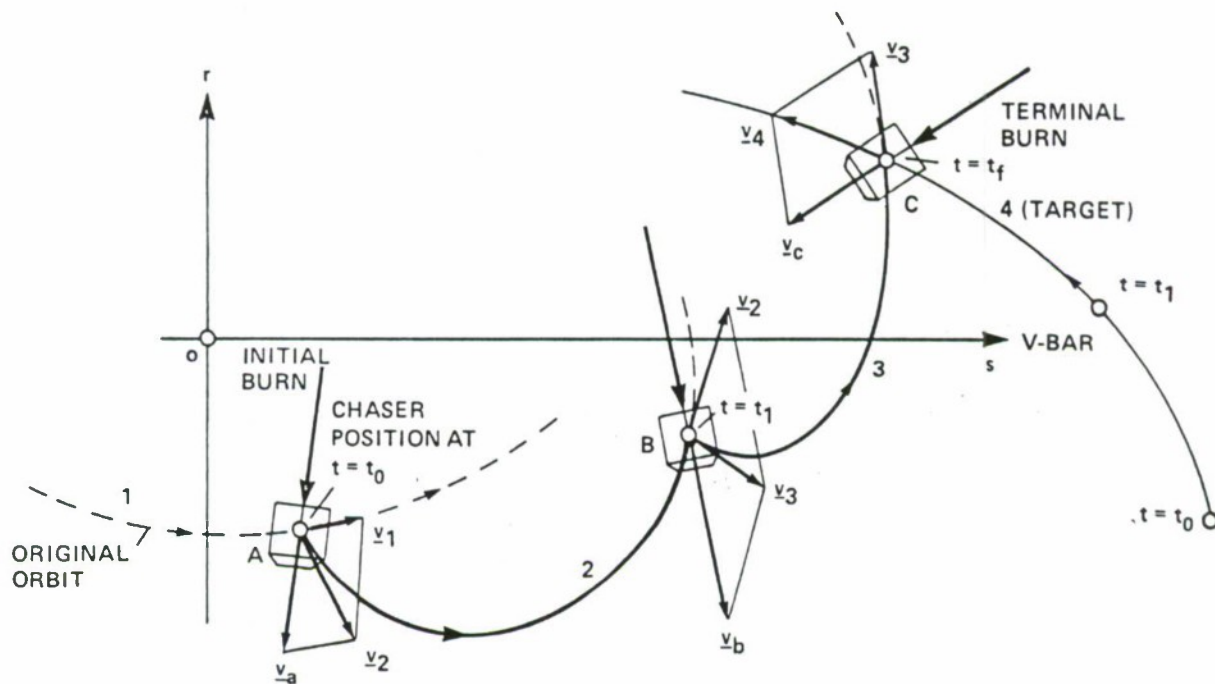
Figure 1.- Example of a three-burn maneuver.

(a)

(b)

(c)

(d)

(e)

(f)

NOTE: BOLD BOX IS ACTIVE WAYPOINT
ENCIRCLED VARIABLES ARE RECOMPUTED
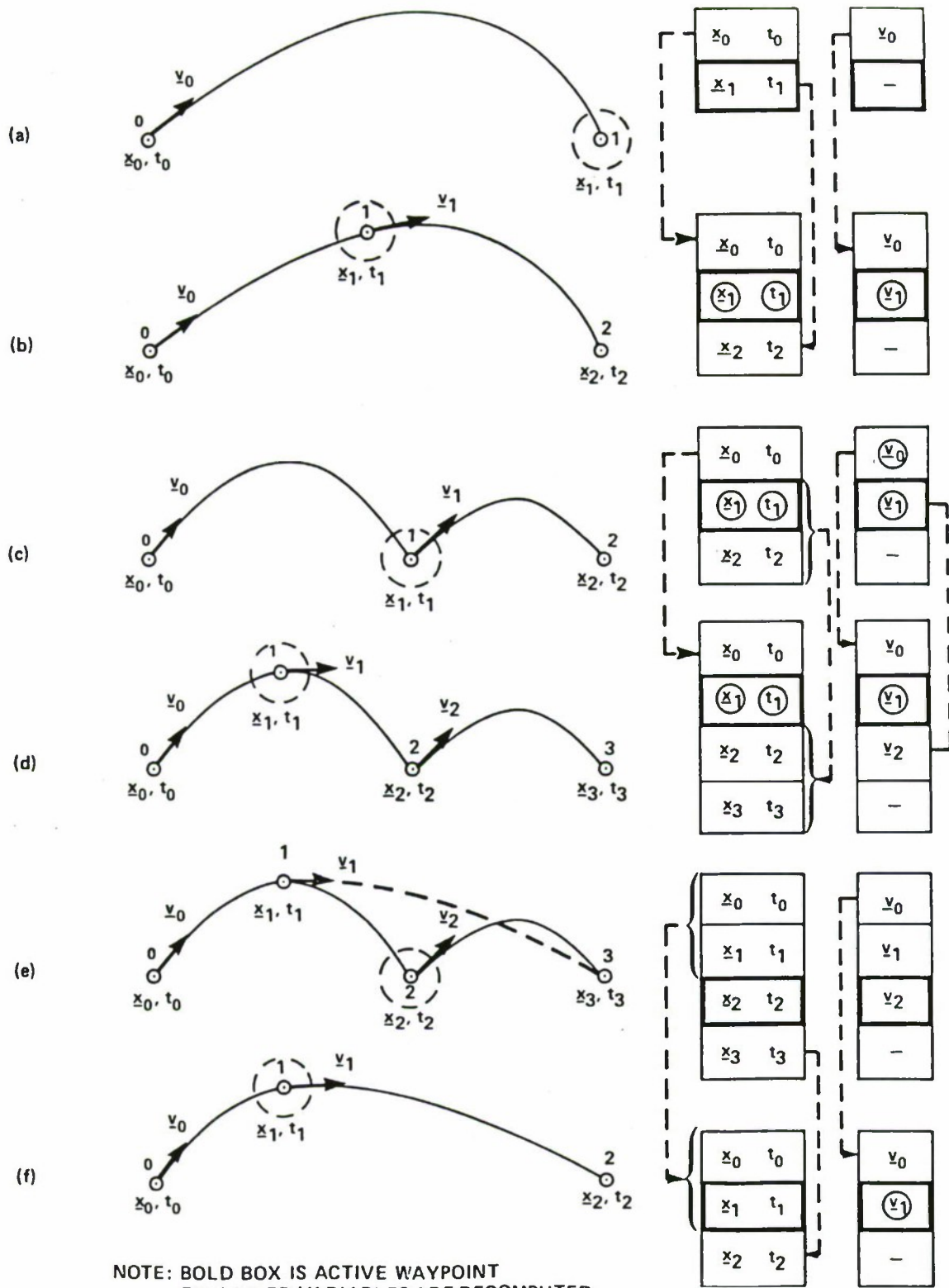
Figure 2.- Editing of way-points.

Figure 3.- Layout of the display area.



Figure 4.- Geometry of the viewing situation.
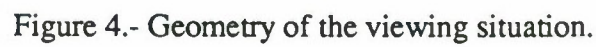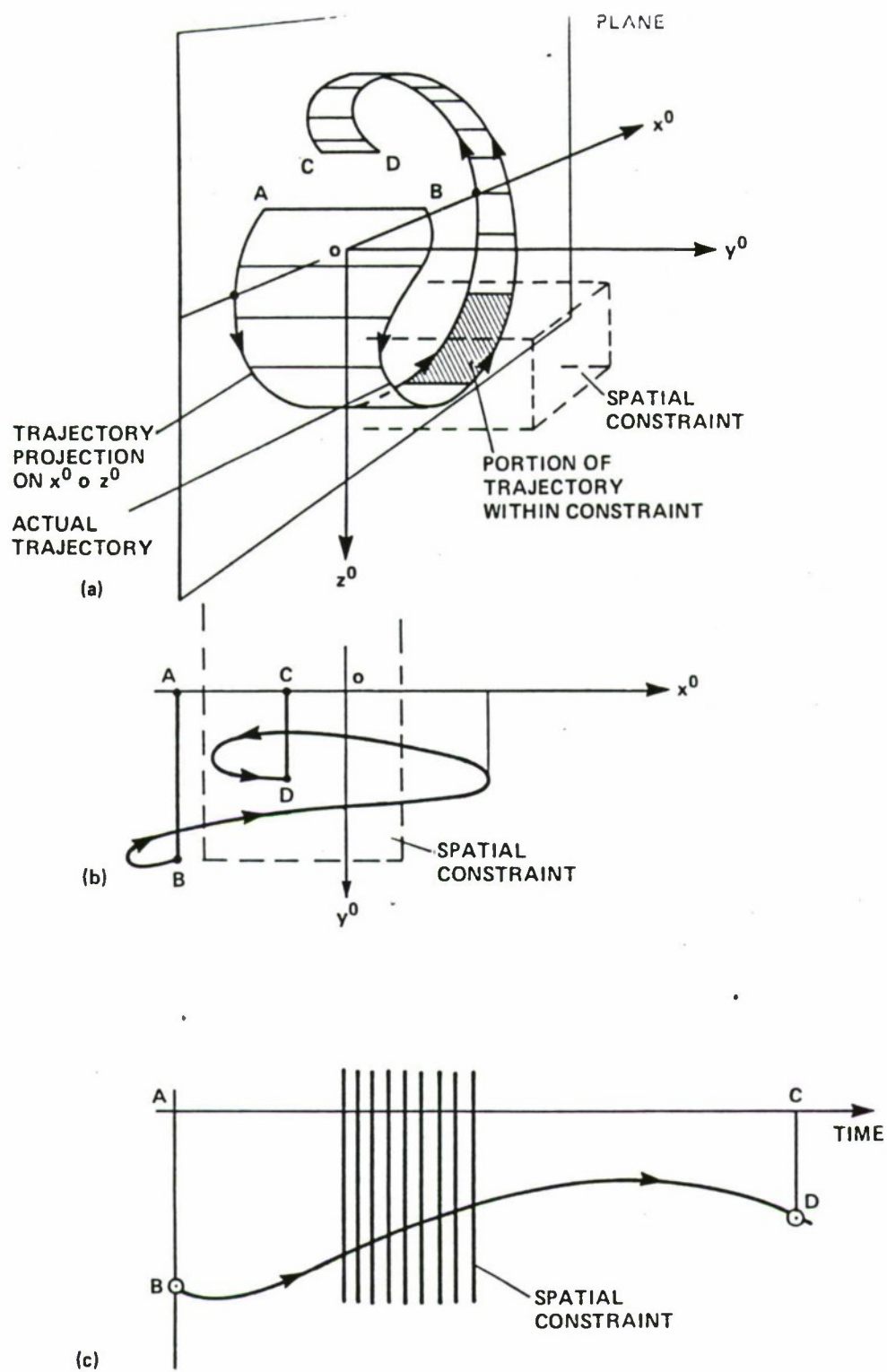
Figure 5.- Principle of the time-axis format. (a) Perspective view. (b) View along the z°-axis. (c) Time-axis format.
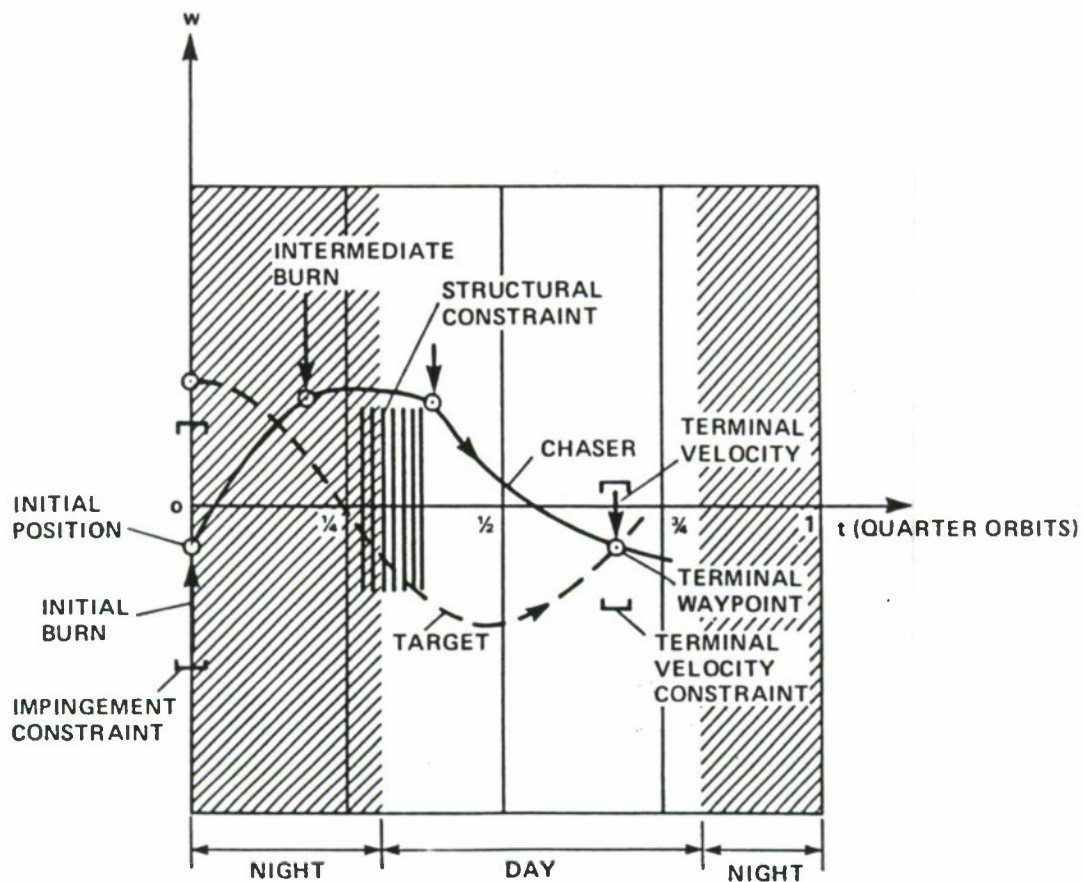
Figure 6.- Time-axis format with operational constraints and day/night indication.
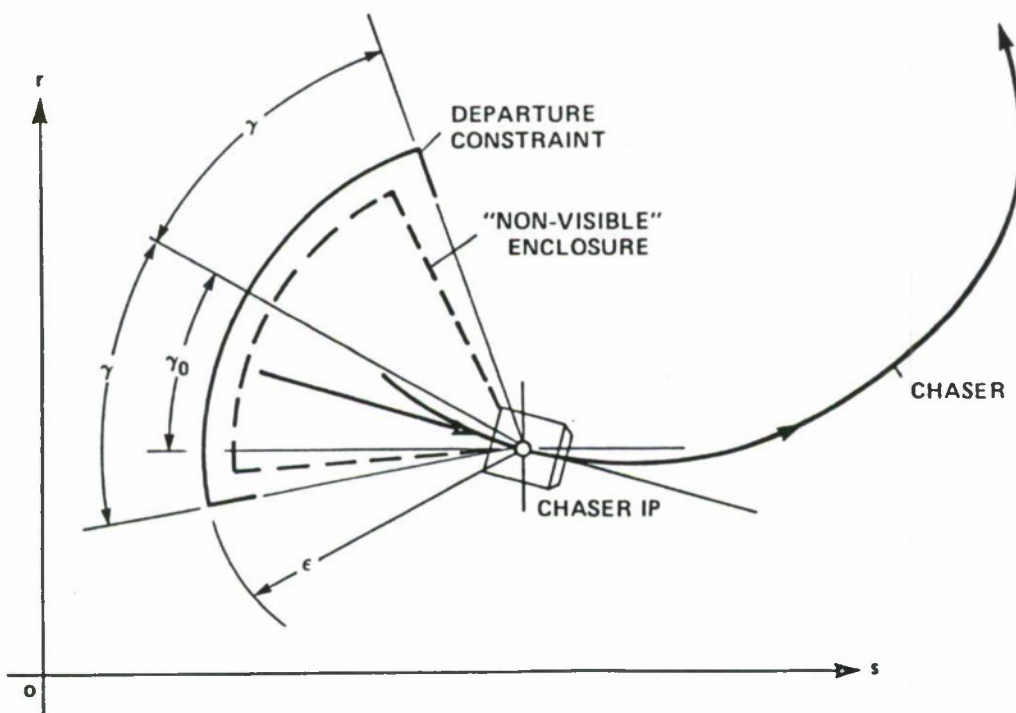


Figure 7.- Departure constraints.
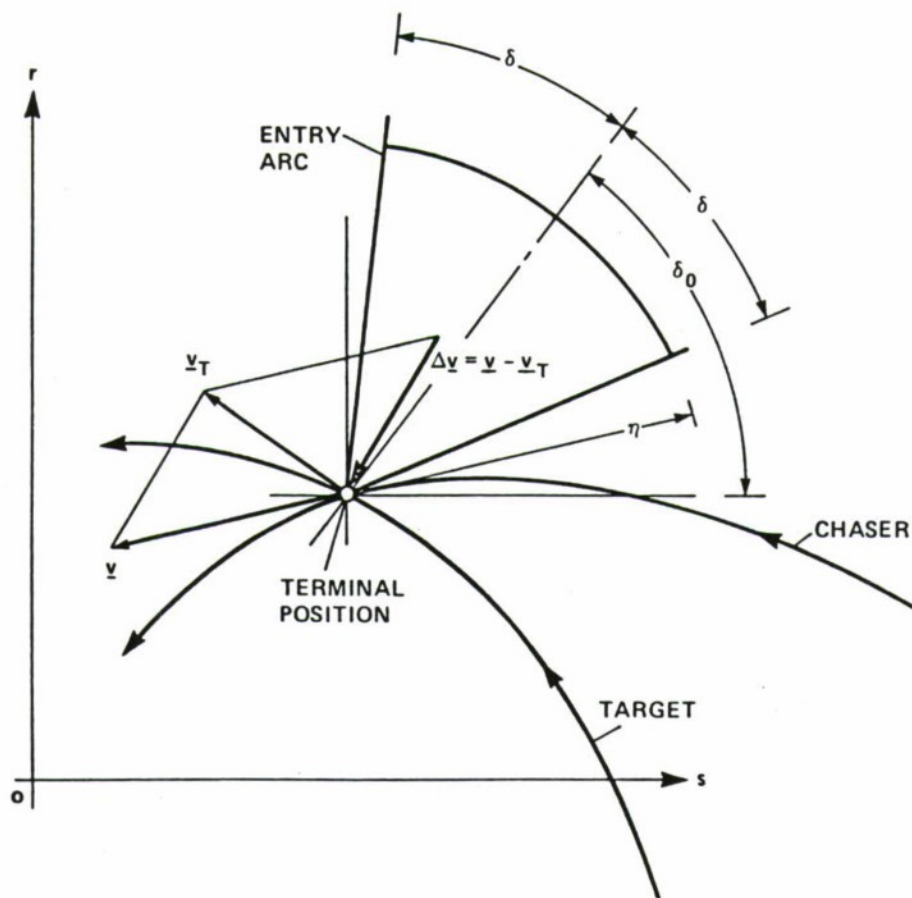
Figure 8.- Arrival constraints.
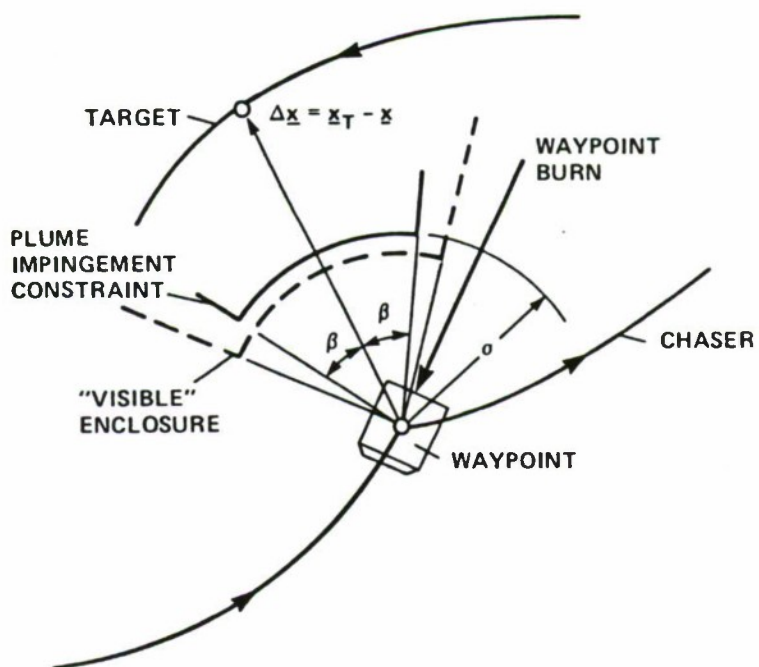


Figure 9.- Plume impingement constraints.
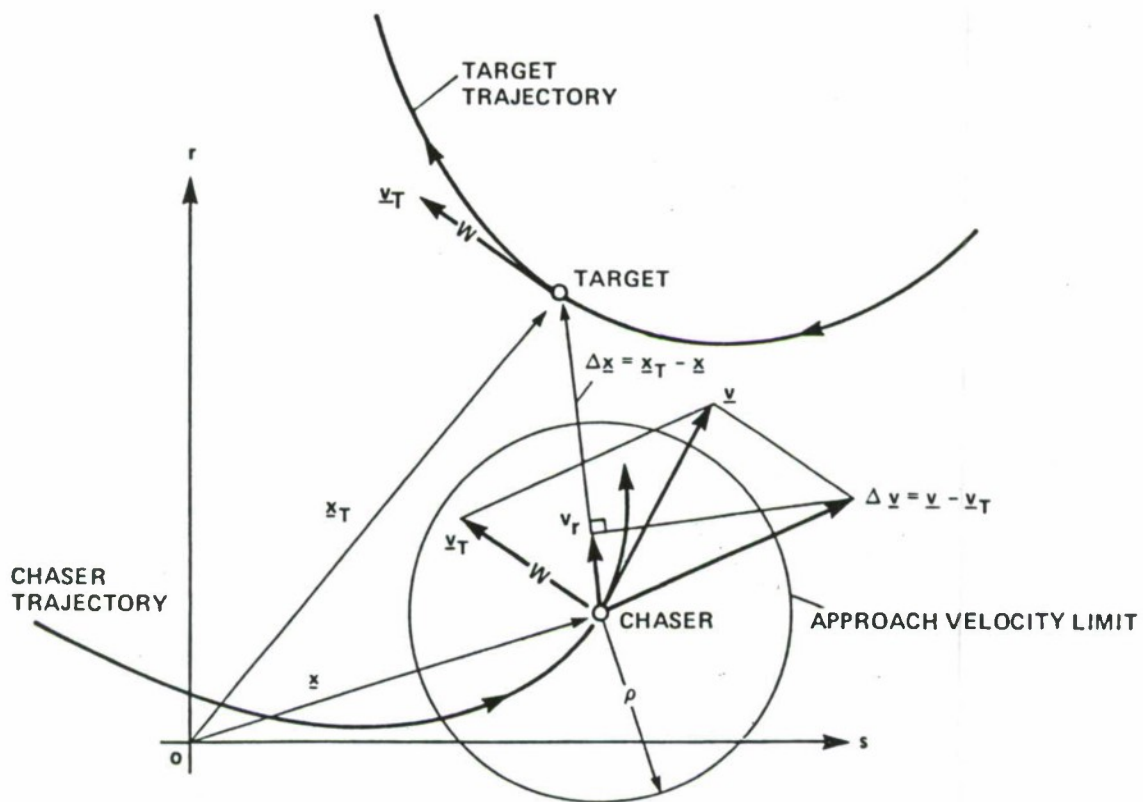
Figure 10.- Approach velocity limit circle.

# EXPERIENCES IN TELEOPERATION OF LAND VEHICLES[1]

Douglas E. McGovern
Advanced Technology Division 5267
Sandia National Laboratories
Albuquerque, New Mexico

## ABSTRACT

Teleoperation of land vehicles allows the removal of the operator from the vehicle to a remote location. This can greatly increase operator safety and comfort in applications such as security patrol or military combat. The cost includes system complexity and reduced system performance. All feedback on vehicle performance and on environmental conditions must pass through sensors, a communications channel, and displays. In particular, this requires vision to be transmitted by closed-circuit television with a consequent degradation of information content. Vehicular teleoperation, as a result, places severe demands on the operator.

Teleoperated land vehicles have been built and tested by many organizations, including Sandia National Laboratories (SNL). The SNL fleet presently includes eight vehicles of varying capability. These vehicles have been operated using different types of controls, displays, and visual systems. Experimentation studying the effects of vision-system characteristics on off-road, remote driving has been performed for conditions of fixed camera versus steering-coupled camera and of color versus black and white video display. Additionally, much experience has been gained through system demonstrations and hardware development trials. This paper discusses the preliminary experimental findings and the results of the accumulated operational experience.

## INTRODUCTION

Remote control of land vehicles can be accomplished through provision of auxiliary sensory channels on-board the vehicle (inside-out control) or through observation of the vehicle in the world (outside-in control). Outside-in control is effective only over short visual ranges for vision with no obscuration by smoke, fog, or obstacles. Inside-out control (referred to as teleoperation in the remainder of this paper) is generally applicable for activities such as security patrols or military combat in which any humans present will be at risk. The cost of such operation is increased complexity in the vehicle and control system, since all knowledge of the environment and the conditions of the vehicle have to be sensed, communicated to a control station, and displayed to the human operator. A further consequence of removng the operator from the vehicle is reduced capability for action, since the information content of the operator feedback is degraded by the intermediary channels.

---

Vehicles, control stations, and teleoperated systems have been built, tested, and demonstrated by a number of organizations. There is little definitive information, however, on the human factors involved in land vehicle teleoperation (ref. 1). Most information has taken the form of a description of vehicle design or proposed application, with only a few papers reporting actual experimental results. Most of the knowledge base is represented by personal experiences and unreported anecdotal evidence. This paper attempts to expand the data base through a presentation of some of the preliminary results of experimentation in teleoperation at Sandia National Laboratories and through discussion of the observations of Sandia personnel gathered over several years of teleoperation experience.

## TELEOPERATION SYSTEMS

Sandia National Laboratories has been actively studying teleoperation for several years. The major effort has entailed the development of a fleet of wheeled vehicles ranging in size from small, interior test beds to large, road and off-road commercial and military vehicles (ref. 2). These vehicles (shown in fig. 1) are being used to conduct feasibility studies on the application of teleoperated vehicles to the physical security and military needs of the U.S. Government. In all of these vehicles, actuators operate the vehicle throttle, brakes, and steering. Control may be derived from manual input at a remote driving station or through some level of automatic control from a digital computer. On-board processing may include simple vehicle control functions or may allow for unmanned, autonomous operation. Communication links are provided for digital communication between control computers, television transmission for vehicle vision, and voice for local control.

Control stations have been developed to support remote operation of the Sandia vehicle fleet. Capabilities range from single television monitor stations with vehicle feedback limited to an audio channel (shown in fig. 2), through large, multiscreen, panoramic displays with computer-generated graphics representations of vehicle speed, pitch, roll, and heading (fig. 3). Vehicle camera mountings have included a single fixed camera, multiple fixed cameras, and cameras slaved to the vehicle steering gear. To date, Sandia has not experimented with stereo vision or with head-slaved displays, although members of the staff have operated such equipment at other locations.

Under the sponsorship of the U.S. Army Missile Command, through the Teleoperated Mobile Antiarmor Platform (TMAP) Project, Sandia has embarked on a major set of experiments to verify some of the observations regarding the "best" driving display (ref. 3). In particular, the experimentation addresses the problems of detection and identification of obstacles in the path of the vehicle. Specific questions include the effect of color versus black and white, the utility of increasing the horizontal field of view through panning a camera in response to steering wheel movements (steering-slaved control), and the errors in operator interpretation of size and distance information as presented by the television system.

## EXPERIENCE

The experimentation on obstacle detection and vehicle control being performed for the TMAP Project represents the only rigorous data base development in process at Sandia. In this

testing, 18 subjects teleoperated a vehicle over a marked off-road course which contained numerous obstacles. An additional 18 subjects participated in a video simulation of the same marked course. Most of the data analysis for this series of tests has been completed (refs. 4 and 5). Additional tests and experimentation are being planned.

The remainder of the experience base at Sandia has been derived from operation of vehicles during hardware and software development and system demonstrations. Operators have ranged from well-trained, highly experienced personnel through people that had not previously driven a remotely controlled vehicle. The primary source of data has been the subjective comments of operators and observers.

The analysis of accidents involving teleoperated vehicles has provided additional information. Table 1 provides a listing. Some of these accidents occurred while the operator was observing the vehicle directly (outside-in operation) and were predominately depth-perception problems involving vehicle clearance or stopping distances. Control reversal caused one accident while operating the vehicle in the outside-in mode. In this accident, the vehicle was heading toward the operator. The operator wanted the vehicle to go toward the left of the operator (operator left). Since the vehicle was approaching the operator, this required the vehicle to turn to the right with respect to its direction of travel. The operator became disoriented and issued a left command. The vehicle responded by veering further to vehicle left (operator right), consequently colliding with a parked car.

TABLE 1.- ACCIDENT HISTORY

| VEHICLE | INCIDENT | CAUSE |
|---------|----------|-------|
| Outside-In Operation | | |
| Dune Buggy | Hit fence | Underestimated stopping distance |
| Dune Buggy | Hit tree | Depth perception |
| Dune Buggy | Hit fence | Underestimated stopping distance |
| Suzuki | Hit post | Depth perception |
| Suzuki | Hit car | Control reversal |
| Inside-Out Operation | | |
| Suzuki | Rollover | Loss of control on hill |
| Suzuki | Rollover | Loss of control on hill |
| Suzuki | Rollover | Hit traffic cone |
| Suzuki | Rollover | Loss of control on hill |
| Suzuki | Rollover | Loss of control on hill |
| Suzuki | Rollover | Loss of control while backing |
| Suzuki | Rollover | Loss of control, hit bump |
| Suzuki | Rollover | Loss of control on hill |
| Suzuki | Rollover | Loss of control, hit bump |

All of the accidents involving teleoperation (inside-out control) have been rollovers. The particular vehicle involved is a small Suzuki LT50 four-wheel, all-terrain vehicle shown in figure 4. The rear wheels are driven through a single-speed drive with a centrifugal clutch. The vehicle is capable of a 15-mph top speed on flat ground. Control inputs from the operator are through the control station illustrated in figure 2. Figure 5 shows the view provided to the operator. In all but one incident, the vehicle was being operated off-road on a motor-cross track with

steep slopes, high banked corners, and high berms at the edges of the track. The only exception was a rollover caused by hitting a traffic cone while operating on a flat asphalt parking lot.

## OBSERVATIONS

A number of observations regarding important parameters, operational considerations, and system design features have been derived from Sandia experiences. These are presented below strictly as indicators since, in the absence of hard experimental data, it is not clear that all are generally applicable. Likewise, not all system implementations are represented.

### Field of View

It is very difficult to operate a vehicle in restricted space with a narrow field of view. Operations of a Jeep Cherokee on normal roads and parking lots were performed with a single camera, 40° field-of-view system. The operator was not comfortable turning corners. Installation of two additional cameras, to provide a total of 120° field of view resulted in much "easier" operation. Additional tests have been run using a steering-slaved camera, both on the Jeep Cherokee and on the Suzuki all-terrain vehicle. Steering-slaved viewing provided sufficient effective field of view to allow turning tight corners and avoiding obstacles. Provision of a mechanism to allow the operator to force the camera further (an auxiliary pan control) was even more effective.

### Resolution

Camera resolution does not seem to be a factor in the ability to teleoperate a vehicle in the absence of obstacles. Sandia has operated vehicles with malfunctioning communications links resulting in extremely poor resolution. As long as operations take place on well defined areas (such as well marked roads) and there are no obstacles in the path of travel, an operator can successfully maneuver a vehicle from one point to another. High resolution does appear to be important when many sizes and types of obstacles are present and for operation off-road where identification of best path is important.

### Color/Black and White

Work with television surveillance systems has indicated that the increased resolution possible with black and white equipment is much more important than any additional information contained in the color signal. This does not necessarily appear true for teleoperation. Color provides additional cues leading to more accurate obstacle recognition and course planning. For example, the difference between dirt and asphalt is important for driving, but cannot be determined from a black and white television picture. Sandia has also found that orange traffic cones (with the color chosen for maximum visibility) tend to disappear on black and white television. These have been used to establish courses during demonstrations and experimentation. Using black and white television, it was found to be necessary to cover the cones with white paper to so that they could be seen.

## Vehicle Vibration

Vehicle vibration and bounce has not been observed to significantly degrade the displayed video scene. The small Suzuki has no suspension (springs or damping) other than its large, soft off-road tires. During operations which lead to the vehicle bouncing enough to actually leave the ground, the video remains relatively clear and usable. No operator has ever commented that vibration or bounce in the picture was bothersome.

## Distance Estimation

As seen from the accident reports, distance estimation during outside-in driving is a problem. It also creates difficulties when using inside-out control. As reported by Spain (ref. 6) in a related set of experiments, operators using a head-mounted display consistently ran into pylons marking the end of a parking place. The feeling of being further from obstacles and landmarks than the actual position has also been reported by most operators of Sandia vehicles. For all of the systems utilized in these observations, however, the display was smaller than geometric similarity, resulting in a scene minification between 0.4 and 0.7. As discussed by Roscoe (ref. 7), it can be anticipated that size and distance judgment errors can be expected for these conditions. To achieve better results, scene magnification of approximately 25% is required.

## Negative Obstacles

Terrain features such as ditches, holes, and drop-offs are extremely difficult to see using television. Negative obstacles such as these have contributed to many of the problems in teleoperating vehicles. In most cases, small ditches cannot be differentiated from variations in ground coloration until the vehicle has hit them. At that point, the horizon on the video scene changes, indicating that the vehicle just hit a ditch. It can be anticipated that stereo vision could help in this problem, but no experimentation has been reported.

## Tilt and Roll

The large number of rollovers reported establish vehicle tilt and roll control as a major problem. In the Suzuki driving system, the only feedback is the video signal from the camera and an audio pickup providing engine sound. Vehicle attitude parameters are neither measured nor displayed. The typical accident scenario entails "launching" the vehicle from a ramp or attempting to traverse a side slope which is too steep for the vehicle to maintain stability. Most rollovers have occurred at close to maximum vehicle speed (about 10-15 mph) and have been a result of ground features representative of extremely challenging terrain. These have included hills with up to 45° slopes and highly banked corners on a motor-cross course. As the rollover occurs, the operators express surprise. In debriefing, it appears that the operator had no indication that the vehicle was approaching a dangerous condition.

## Overcontrol

A typical characteristic of novice operators is extreme steering overcontrol. The operator applies a small steering input to the vehicle, but no result is immediately seen. The steering input is increased until a response is finally observed. The resulting turn is more than intended so the operator applies a small correction. Again, the response is not seen so more correction is applied, etc. The outcome is vehicle travel oscillating about the desired path. Operators report this to be a very stressful situation. Overcontrol has also contributed to several of the vehicle rollover accidents. The operator applied excessive steering input, sending the vehicle over the edge of a berm. Observing novice drivers learning to control the vehicle, it is apparent that considerable internal control is being exercised as the operator adapts. After some minutes of operation, steering operation is considerably slower and at lower amplitude, resulting in smoother vehicle control. Spain (ref. 6) reports similar findings.

## Navigation

An associated problem in vehicle teleoperation is the difficulty of maintaining spatial orientation with respect to major landmarks, map features, or compass directions. It is not uncommon for operators to become lost on the motor-cross course. Even with landmarks and a map of the course, they have not been able to determine how to return to the starting location without assistance.

## SUMMARY AND CONCLUSIONS

Operational experience has been gathered at Sandia through development, test, and demonstration of a number of vehicles. A large experimental program in vision system requirements for teleoperation is also in process. Through the knowledge gained in these programs, several key areas can be identified as critical to successful control of a teleoperated vehicle. The primary area is the quality of the visual display provided to the operator. It has been shown that vehicles can be controlled in restricted environments with extremely poor conditions of viewing. As viewing improves (both in resolution and field of view), better control can be expected.

Negative obstacles create difficulty in that operators cannot distinguish them from other terrain features which do not affect vehicle travel. The result is hitting ditches, holes, or berms at excessive speed.

The interaction of the vehicle with the environment, as interpreted through the mediating effects of the television display system, can lead to poor control capabilities and hazardous operating conditions. Overcontrol of the vehicle steering, coupled with the operator's inability to accurately perceive vehicle attitude and terrain requirements has led to a number of accidents. This can be partially linked with the absence of kinesthetic feedback to the operator. Experimentation with vehicle simulators has shown a distinct lag in response to environmental inputs, such as wind gusts, when no kinesthetic feedback is present (ref. 8). With the addition of kinesthetic feedback to the operator (simulator platform motion), response time to sudden wind gusts dropped from an average of 0.56 sec to an average of 0.44 sec. Similar results have been reported for the addition of steering wheel torque feedback, thus providing "feel of the road" to the operator (ref. 9). The

lack of kinesthetic feedback is similar to operating with a time delay in the control system. Additional lags are introduced by the communications systems and vehicle actuator and control systems.

Given the ability to maneuver a teleoperated vehicle in the real-world environment, the problem of navigation is encountered. Operators tend to get lost, disoriented, and confused when provided with visual input and maps. The effect of addition of vehicle heading, plotting of route traveled, or other aids remains to be investigated.

# REFERENCES

1. D. E. McGovern, "Technology Assessment for Command and Control of Teleoperated Vehicles", SAND87-1225, Sandia National Laboratories, Albuquerque, May 1987.

2. D. E. McGovern, "Advanced Technology Mobile Robotics Vehicle Fleet", SAND87-0033, Sandia National Laboratories, Albuquerque, March 1987.

3. D. P. Miller, "Evaluation of Vision Systems for Teleoperated Land Vehicles", 1987 IEEE Systems Man and Cybernetics Conference, October 20-23, 1987, Alexandria, VA.

4. D. P. Miller, "Evaluation of Vision Systems for Teleoperated Land Vehicles," IEEE Contr. Sys., Vol. 8, No. 3, pp. 37-41, 1988.

5. D. P. Miller and D. E. McGovern, "A Laboratory-Simulation Approach to the Evaluation of Vision Systems for Teleoperated Vehicles," prepared for presentation at the International Symposium on Teleoperation and Control, University of Bristol, England, 12-15 July 1988.

6. E. H. Spain, "Assessments of Maneuverability with the Teleoperated Vehicle (TOV)", Fourteenth Annual Symposium of the Association for Unmanned Vehicle Systems, July 19-21, 1987, Washington, DC.

7. S. N. Roscoe, "When Day is Done and Shadows Fall, We Miss the Airport Most of All", Human Factors, Vol. 21, No. 6, pp. 721-731, 1979.

8. W. W. Wierwille, J. G. Casali, and B. P. Repa, "Driver Steering Reaction Time to Abrupt-Onset Crosswinds, as Measured in a Moving-Base Driving Simulator", Human Factors, Vol. 25, No. 1, pp.103-116, 1983.

9. R. W. Allen and D. H. Weir, "Analysis of Man-in-the-Loop Performance Measurement Technology for Crash Avoidance Research", STI TR-1200-1, Systems Technology, Inc., Hawthorne, CA, March 1984.

Figure 1.– Teleoperated vehicles.
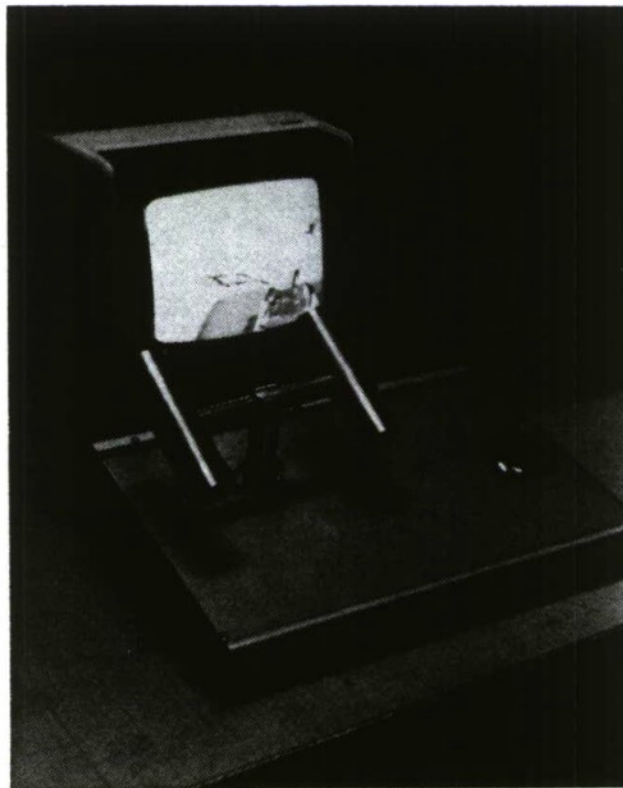


Figure 2.– Single monitor with audio feedback.

Figure 3.– Panoramic display.

Figure 4.– All-terrain vehicle.



Figure 5.– Operator's view via control station.

# DEVELOPMENT OF A STEREO 3-D PICTORIAL PRIMARY FLIGHT DISPLAY

Mark Nataupsky[1]
NASA Langley Research Center
Hampton, Virginia

Timothy L. Turner
Harold Lane
Lucille Crittenden
Research Triangle Institute, North Carolina

## INTRODUCTION

Computer-generated displays are becoming increasingly popular in aerospace applications. The use of stereo 3-D technology provides an opportunity to present depth perceptions which otherwise might be lacking. In addition, the third dimension could also be used as an additional dimension along which information can be encoded.

Historically, the stereo 3-D displays have been used in entertainment, in experimental facilities, and in the handling of hazardous waste. In the last example, the source of the stereo images generally has been remotely controlled television camera pairs.

This paper describes the development of a stereo 3-D pictorial primary flight display used in a flight-simulation environment. The purpose of this research is to investigate the applicability of stereo 3-D displays for aerospace crew stations to meet the anticipated needs of the 2000-2020 time frame. Although the actual equipment that could be used in an aerospace vehicle is not currently available, the laboratory research is necessary to determine where stereo 3-D enhances the display of information and how the displays should be formatted.

## HARDWARE/SOFTWARE CONFIGURATION

The hardware consists of a VAX 11/780 computer, an Adage 3000 raster programmable display generator (PDG), and a Stereographics 3-D display stereoscopic system. A FORTRAN aircraft simulation is used to provide parameters to the display programs residing in the Adage 3000. The display programs are written in a "C" language known as ICROSS-3000, with a graphics-enhancement package known as the Real-Time Animation Package (RAP). (RAP is a proprietary software product developed at the Research Triangle Institute.)

The Stereographics display uses liquid crystal shuttered glasses and specially adapted hardware which divides each video frame into two fields corresponding to the left- and right-eye views, each at half the resolution. The PDG outputs a 60-Hz repeat field, 512 x 512 pixel image. The stereo display system converts this input to a 120-Hz repeat field, 216 x 512 pixel output with

---

[1]Lt. Col., USAF. Detailed to the NASA Langley Research Center.

alternating left- and right-eye fields. Figures 1 and 2 show a monocular version of the display. Figure 3 shows a similar display with left- and right-eye stereo views as they would appear on a conventional 60-Hz monitor. The Stereographics system converts the input shown in figure 3 and generates the stereo pairs similar to those in figures 1 and 2, but with only half the vertical resolution. The liquid crystal shuttered glasses are synchronized so that each eye sees only one of the stereo views.

A stereo image pair contains twice the information contained in a monocular image. Therefore, on a system with limited video bandwidth, either the video frame rate or the number of lines must be reduced when stereo displays are being generated. The current system maintains frame rate by halving the number of lines. Flicker, which was a problem with other systems, is thus eliminated. The system also performs the conversion of the video signal, and the PDG responds as if it were outputting its customary 60-Hz repeat field image. The liquid crystal shutter technology is much faster than the video frame-rate-display capabilities; therefore, the stereo system does not impose any bandwidth limitations.

## DISPLAY FEATURES

The main features of the display are an own-ship symbol, a perspective follow-me target ship, two different 3-D tracks showing the path of the target ship, a ground grid around the runway, a pitch grid on both the left and right sides of the display, and digital readouts for altitude/heading/airspeed. The digital readouts display the instantaneous values for the own-ship and the desired preprogrammed flightpath. Because the own-ship remains fixed relative to the other display elements, an inside-out (i.e., moving horizon) display is represented.

### Generating the Stereo 3-D Effect

The display program needs to generate the left- and right-eye views of the display. Given distinct x, y locations of each eye, the calculation of the viewing transformations are described by Foley and Van Dam (ref. 3).

Two parameters are used to control the stereo 3-D effect: zero-parallax distance and interocular separation. In general, parallax refers to the positional discrepancy in the left- and right-eye views of a point in the display. The parallax is zero when the corresponding points in each view occupy the same relative screen location. Points in the display at the zero-parallax distance from the eye appear to lie in the plane of the screen. Points closer to the pilot than the zero-parallax distance appear to lie in front of the screen, while points farther from the pilot appear to lie beyond the screen. In addition, the interocular distance controls the apparent relative depth of objects in the display. The greater the interocular distance, the more powerful the stereopsis effect. By comparing the apparent depth of the target ship with the own-ship symbol, the pilot has an indication of position error. This stereo 3-D effect reinforces the depth cue provided by the relative size of the perspective target ship.

When viewing objects in the natural environment, the eyes must perform the separate functions of converging and focusing on a point of interest. In a stereoscopic display, although the eyes must converge on an object, they focus on the plane of the screen regardless of the apparent

distance. One requirement of a stereo 3-D display is to minimize that disparity (ref. 4). This is accomplished by keeping the principal objects near the zero-parallax distance where the focus and convergence relationship is correct.

After setting the zero-parallax distance at the desired distance from the aircraft to the target ship, the size of the target ship becomes a distance cue. For example, if the pilot is following too closely, the target ship appears larger on the screen and projects out of the plane of the screen towards the pilot. Conversely, if the pilot drops behind the target ship, it appears to shrink in size and recede into the background. The combination of stereo and size cue serves as an important error indicator.

In this display, the zero-parallax distance is set to a nominal following distance. The interocular distance was established empirically at 8 ft. Moving the eyes that far apart is equivalent to shrinking the scale of the scene proportionally. Such distortions enhance the pilot's ability to perceive the sensations of depth. They are also necessary because of inherent limitations of the hardware. The precision in rendering the left- and right-eye views is limited both by the display resolution and the arithmetic precision of the display processor (i.e., 16-bit fixed point).

If a fixed time lag rather than a fixed distance is desired, the zero-parallax also could be dynamic. In that case, the zero-parallax distance would be a function of the time lag and the instantaneous velocity of the target ship.

Within the 3-D display, apparent depth had to be assigned to 2-D symbols such as digital readouts and the pitch scale. Two possible choices are the zero-parallax distance or the maximum distance. If they are set at the zero-parallax distance (i.e., drawn with the same left- and right-eye view), they would be perceived by the pilot as if they were being looked past in order to see the part of the 3-D display beyond the zero-parallax distance. Earlier informal evaluation has shown the resulting perception to be disorienting. Instead, by placing the 2-D symbols at the maximum distance, they appear natural and unobtrusive.

Care must be taken when defining the left- and right-eye transformations. Figure 4 illustrates two ways of conceptualizing the transformations. In figure 4a, the views are converged by rotation of the viewing pyramid. In figure 4b the viewing pyramids are sheared. The latter approach is preferred, as the projection planes in each view remain parallel. Achieving convergence by rotation creates artifacts which can not only cause eye fatigue, but also can interfere with the pilot's perception of depth (ref. 4). Although the rotation method is easier to implement, the shearing approach has become the standard in 3-D graphics software (refs. 1 and 2).

## Own-Ship Symbol

Figure 5 shows the evolution of the own-ship symbol. The original configuration, figure 5a, presented the pilot with two problems. First, it was impossible to perceptually fuse the right- and left-eye viewpoints to form the 3-D image. This fusion problem was surprising, because the signposts also were made of single, straight lines, but there was no problem with their visual fusion. An additional problem was that the own-ship symbol tended to "get lost" in the display. The signpost symbol was constructed of perpendicular horizontal and vertical lines; the same was true of the own-ship symbol. Therefore, there were many instances in which the own-ship symbol would overlay the signposts and could not be perceived.

In order to increase the pilot's ability to perceive the own-ship symbol, the center slanted lines were drawn as shown in figure 5b. Although the ability to perceive the symbol was greatly increased, there was still the problem of inability to visually fuse the stereo 3-D image.

Figure 5c was originally constructed to further enhance the pilot's ability to perceive the own-ship symbol; it worked. A serendipitous benefit was that the symbol now visually fused. At this time there is no theoretical explanation for the fusion phenomena.

## INITIAL RESEARCH

The initial research with the display will be a study of recovery from flightpath offset. Pilots will be initiated on the nominal flightpath. After 2 sec, they suddenly will experience a flightpath offset. They will be required to make the stick input to rejoin the nominal flightpath. Visually evoked potentials will be triggered from the sudden flightpath offset. In addition, reaction times, response accuracy, and a projected workload estimate also will be recorded. The Subjective Workload Assessment Technique (SWAT) will be used for the workload estimate (refs. 5 and 6). A test for stereoscopic acuity will be administered prior to data collection. Recent anecdotal evidence indicated that some subjects tend to lose the ability to use the stereoscopic cue after prolonged exposure to it. Therefore, stereoscopic acuity also will be measured immediately after a long series of trials with the stereo 3-D cues.

In addition to using stereo 3-D or monocular cues as an independent variable, the inclusion or exclusion of the target ship will be the second independent variable. The last independent variable will be the pathway. There either will be the signpost or a monorail pathway for the subjects to follow.

## FUTURE RESEARCH

The initial research will use the stereo 3-D cues to represent geographic information. In the "real world," objects are geographically separated by space, and the displays will attempt to create the perception of that geographic separation.

In contrast, one line of future research will use the third dimension as a dimension to encode new information for the pilot. For example, presume that there is a pictorial display which is entirely in the plane of the screen and that depth perception is simulated with monocular cues such as linear perspective. If a pilot were using that display in a current aircraft, and if the airspeed were to get too low, an audio display (i.e., a horn) would sound. The audio display is an alerting display, and the pilot must know to then look at the visual display for speed.

However, part of the pictorial display is a box with digital readouts for instantaneous actual and desired airspeed. Using the same airspeed error example, the box with the airspeed would modulate in the third dimension (i.e., along the z-axis) as the alerting cue instead of using the audio cue as the alerting cue. In this manner, new information would be presented to the subjects in the third dimension.

From a human factors perspective, that is a potential way of decreasing the total number of cockpit displays and also to make the alerting cues more nearly intuitively obvious. There are many research questions to be addressed. First, can it be demonstrated that the proposed use of stereo 3-D is quantifiably better than the use of audio alerting cues? Some of the other questions concern the rate and perceived depth of modulation in the third dimension. For example, should the rate or perceived depth of modulation be proportional to the amount of error? Should the modulation only be from the plane of the screen towards the pilot or should it also modulate from the plane of the screen away from the pilot?

Other uses of stereo 3-D also are possible. The "natural" use of stereo 3-D is to represent the 3-D geography. Part of the true test of the technology will be to go beyond that approach and determine if there are more effective applications.

# REFERENCES

1. American National Standards Institute. Draft Proposed American National Standard Programmer's Hierarchical Interactive Graphics System. Prepared by Technical Committee X3H3 - Computer Graphics. New York: ANSI, 1985.

2. Association for Computing Machinery. Status Report of the Graphics Standards Planning Committee. Computer Graphics, Vol. 13, No. 3, Aug. 1979.

3. Foley, James D. and Van Dam, Andries. Fundamentals of Interactive Computer Graphics. Reading, MA: Addison-Wesley, 1982. Reprinted with corrections, July 1984.

4. 3-D Display Handbook for Stereoscopic Computer Graphics. San Rafael: Stereographics Corporation, 1984.

5. Reid, Gary B., Shingledecker, Clark A., and Eggemeier, F. Thomas. Application of Conjoint Measurement to Workload. Proceedings of the Human Factors Society-25th Annual Meeting. Santa Monica, CA: Human Factors Society. 1981a.

6. Reid, Gary B., Shingledecker, Clark A., and Eggemeier, F. Thomas. Development of Multidimensional Subjective Measures of Workload. Proceedings of the International Conference on Cybernetics and Society, IEEE Systems, Man and Cybernetics Society. 1981b.
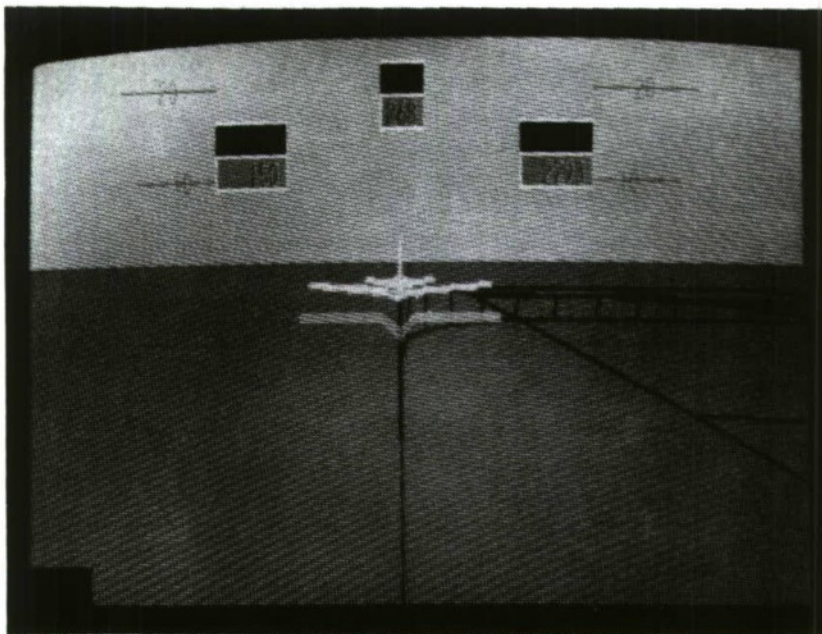
Figure 1.- Monocular "monorail" display.
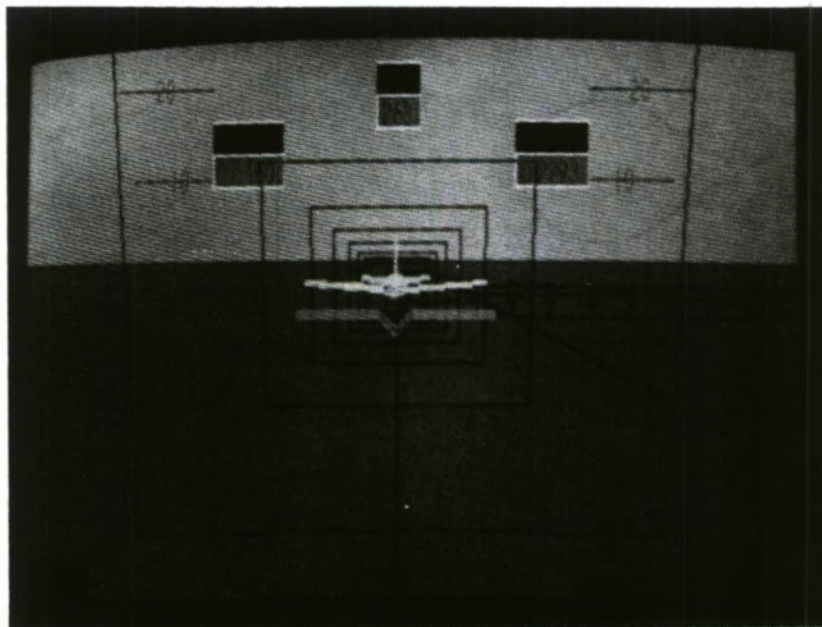


Figure 2.- Monocular "signpost" display.
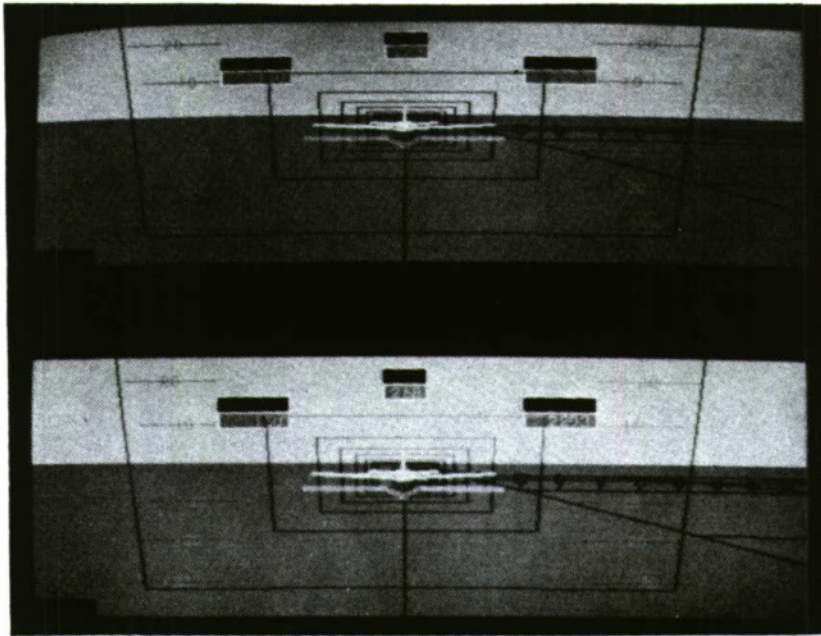
Figure 3.- Stereo display as seen on a conventional CRT.
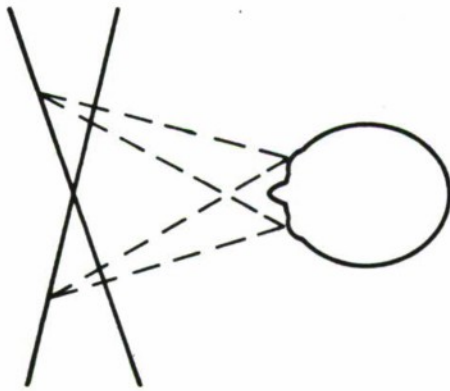


(a)
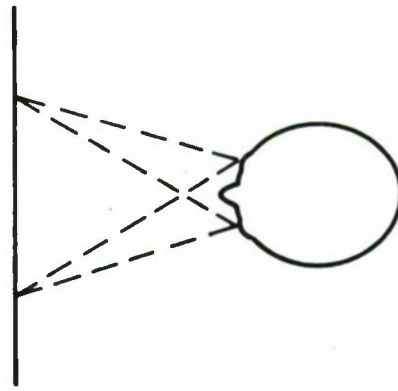
(b)

Figure 4.- Generation of stereo pairs by eye rotation (a); generation of pairs by shearing the viewing pyramid (b).



(a)

(b)

(c)

Figure 5.- Evolution of own-ship symbol:  Stereo pairs for (a) and (b) would not visually fuse; (c) would visually fuse.

# SYNTHETIC PERSPECTIVE OPTICAL FLOW: INFLUENCE ON PILOT CONTROL TASKS

C. Thomas Bennett, Walter W. Johnson, John A. Perrone
NASA Ames Research Center
Moffett Field, California

Anil V. Phatak
Analytical Mechanics Associates
Palo Alto, California

## INTRODUCTION

Computational and empirical analyses of optical flow have led to a more complete understanding of pilot control tasks. Such analyses are based on the premise that a primary stimulus for the perception of self-motion is the flow of optical texture in the visual field (Gibson, 1950; Koenderink and van Doorn, 1976). It has been further recognized that there are both local and global optical variables that might influence control behavior (Owen and Warren, 1982; Uttal, 1985). With this realization came the understanding that to study how optical flow influences control tasks, it is essential that the complex visual scene be decomposed into observable flow patterns (Regan and Beverly, 1985).

One approach used to better understand the impact of visual flow on control tasks has been to use synthetic perspective flow patterns. Such patterns are the result of apparent motion across a grid or random dot display. Unfortunately, the optical flow so generated is based on a subset of the flow information that exists in the real world. The danger is that the resulting optical motions may not generate the visual flow patterns useful for actual flight control.

We have conducted a series of studies directed at understanding the characteristics of synthetic perspective flow that support various pilot tasks. In the first of these, we examined the control of altitude over various perspective grid textures (Johnson et al., 1987). Another set of studies has been directed at studying the head tracking of targets moving in a three-dimensional coordinate system. These studies, parametric in nature, have utilized both impoverished and complex virtual worlds represented by simple perspective grids at one extreme, and computer-generated terrain at the other.

These studies are part of an applied visual research program directed at understanding the design principles required for the development of instruments displaying spatial orientation information. The experiments also highlight the need for modeling the impact of spatial displays on pilot control tasks.

# ALTITUDE CONTROL

## Introduction

The purpose of this experiment was to examine the characteristics of "wire frame" perspective grids as support for altitude control. Wolpert, Owen, and Warren (1983) reported that splay angle information was one of the most important indicants of altitude change. In their study, they used ground surface textures consisting of equally spaced lines either parallel to the direction of travel (meridian texture), orthogonal to the direction of travel (latitudinal texture), or both (square texture).

There are two limitations of Wolpert's work that have relevance to the current study. The first is that discrete-trial, passive-response methodology was used. This is in contrast with a setting where a person is required to continuously monitor a perspective scene, and where his or her responses result in feedback control of perspective dimensions of the stimulus.

The second limitation derives from the fact that subjects could have monitored the location at which any meridian texture line intersected the bottom edge of the screen. As a result, a subject could tell if altitude had changed by merely observing the movement and intersection without monitoring the splay angle at all.

## Methods

Subjects were flown at a constant velocity, at three different altitudes, over each of the three grid types mentioned above. The display was generated by an Evans and Sutherland PS-2 graphics system. The "aircraft" was buffeted by both lateral and vertical winds. Each of the disturbances was defined by its own sum of 13 sine waves. The five subjects were required to maintain a constant height above the grid by means of a joy stick. The primary performance metric was adjusted root mean square error (ARMSE) from the assigned altitude.

The important point here is that because of the lateral noise imposed on the craft position, the meridian lines moved left and right irrespective of the actual change in altitude. As a result, subjects could not determine altitude change by only the movement of the meridian lines. Changes in altitude would have to be determined by changes in density (lower density corresponds to a lower altitude) and splay angles (the greater the angle the lower the altitude) of the grid structure.

## Results and Discussion

Based on the work previously cited, it was expected that ARMSE sould be lowest for the meridian surface and highest for the latitude surface. This was not the case (fig. 1).

Because of the unexpected larger ARMSE values obtained when flying over the meridian surface texture, it was decided to look more critically at a single subject's performance. A detailed power frequency analysis was performed and showed that the meridian grid resulted in (1) less stick power associated with the vertical disturbance than any of the other grid textures; and (2) the most power in the stick movement associated with the lateral input signal (fig. 2).

These analyses indicate that the subject (1) was less reactive to the information specifying true changes in altitude when flying over the meridian texture; and (2) tended to confuse lateral with vertical motion in displays where only splay information was available.

# PERSPECTIVE FLOW FIELDS AND HEAD TRACKING IN A 3-D VIRTUAL WORLD

## Introduction

In the previous study, we discussed the impact of perspective flow displays on a manual control task that regulated the altitude of a simulated aircraft. In certain military rotorcraft, systems exist in which movement of a sensor system is slewed to the crewmember's head motion. Currently there is only standard flight symbology in this helmet-mounted display to indicate altitude, attitude, and heading. A small portion of the display provides information concerning the field of view and field of regard of the sensor.

Despite the fact that these systems are currently fielded, little systematic data exist concerning how a pilot uses flight/target information presented on a helmet-mounted display. Even less data are available on alternative display configurations that might make a pilot more sensitive to changes in aircraft state.

As part of a program to better understand helmet-mounted flight displays, we conducted a study to validate a laboratory simulation of the currently fielded system. A perspective flow field was used to create the virtual world that was the basis for this simulation. A detailed report of this study is in preparation.

## Methods

A wire-frame perspective grid was displayed to six subjects by means of a head-mounted 1 in. Sony electronic viewfinder. Head position was monitored by means of a Polhemus head tracker. As the subjects moved their heads, they were able to "look" around the virtual world.

Six subjects were "flown" over the grid at two different altitudes and three different velocities. Positioned on the surface was a wire frame cube. The target was offset to the left or right of the direction of travel. The subject could "track" the target by means of a cross hair that was generated in the middle of the monocular display. Tracking ARMSE was determined by subtracting line of sight (LOS) to target from the visual LOS.

## Results and Discussion

Figure 3 shows the mean screen errors for the different offsets, as a function of slant angle to the target. The term slant angle incorporates elevation and azimuth components. It is important to remember here that as range to the target decreases, optical (apparent) velocity increases. So, during the course of the "flight," the target was in fact accelerating, even though "aircraft" speed

was constant throughout the flight. A 3×4×2 (speed × offset × altitude) repeated-measures analysis of variance was conducted on the mean ARMSE values for each subject. This analysis indicated that as optical velocity increased, there was a significant increase in screen error ($p < 0.001$). This was true irrespective of whether the increase in optical velocity was produced by changes in slant range or "vehicle" speed.

In figure 4 is shown the change in both ground error and screen error as a function of slant angle. To calculate ground error, the target and visual LOSs were first projected to the ground plane. Ground error was then given as the distance between those two intersections. As slant angle increased, ground error did not significantly change ($p > 0.46$). One interpretation of these data is that the subjects were treating the task as a true three-dimensional LOS problem. If the subjects had maintained screen error constant (as in an arcade game), ground error would have directly varied with slant range. A second interpretation is that subjects tried to maintain a constant screen error, but were unable to do so because of the accelerating optical velocity of the target.

# HEAD TRACKING DURING SIMULATED AUTOMATED AND MANUAL HELICOPTER FLIGHT

## Introduction

A model of head tracking in a 3-D world (represented by a perspective flow field) was developed and tested in the previous study. The purpose of the present experiment was to (1) validate the laboratory simulation, and (2) model the trade-offs that pilots make when they are required to control their craft and simultaneously head-track targets. A detailed report of this study is in preparation.

## Methods

Six AH-64 Apache helicopter pilots took part in a simulation of the pilot night-vision system (PNVS). The study took place in a fixed-base mock-up of the helicopter. The visual scene was a complex, computer-generated world in which a stationary helicopter served as the target. Each pilot was initially flown "automatically" in either a rectilinear or curvilinear path past the target. This served to simulate a copilot/gunner or a pilot in an automated flight mode. The pilot was then required to duplicate the ground track in manual flight mode while simultaneously tracking the target. The spread of target ranges extended from approximately 6,000 to 400 ft. In the trials reported here, own-ship velocities never exceeded 80 mph.

Head-tracking ARMSE was calculated as in the previous study. Ground-track error was also measured. This was the difference in feet between the flightpaths in the automated versus manual flight modes. During the manual flights, pilots were informed that target tracking was the primary task, but that ground track error was being measured.

# Results and Discussion

Figure 5a shows the averaged screen errors in the manual and automatic flight modes, as a function of slant angle. A repeated-measures analysis of variance revealed a significant effect of slant angle ($p < 0.005$) as well as significant slant angle by flight mode interaction ($p < 0.001$). The inference is that screen error is greater near the end of a manual flight than it is at the end of an automatic flight.

At first glance this makes a great deal of intuitive sense. During manual flight, the pilot is not only head-tracking a target, but also manually flying the helicopter. However, inspection of figure 5b reveals another explanation of the increased screen error. As can be seen, optical velocities during the manual flight mode are significantly greater than during automated flight. Additionally, a multivariate regression revealed a significant positive correlation ($p < 0.0001$) between optical velocity and screen error, when the effect of slant angle is statistically removed. This analysis is consistent with the interpretation that optical velocity is a major source of head-tracking error.

An interesting question that arises from these data is why optical velocities are greater during manual flight. Presumably, given that the pilot is under control of the craft, he or she could have biased the flightpath to decrease optical velocity, and, hence, screen error.

Figure 5c provides some understanding of the complex trade-offs that the pilots were making. This figure shows that as slant angle increased (and slant range decreased), the magnitude of the ground error decreased significantly ($p < 0.005$), then gradually increased. As with the second experiment, the data reported here are consistent with the interpretation that the pilots were treating the task as a true 3-D problem. Otherwise, there would have been no reason why they would not have simply held screen error constant and allowed ground error to vary. Also, although they flew a flightpath that increased the problem of head tracking (by increasing optical velocity), their manual flightpath resulted in, if not a constant, at least a minimal ground error. This, of course, is the name of the game for a combat pilot.

# GENERAL DISCUSSION

Pilot control tasks include both manual flight control and the control of head-slaved sensor systems. Three studies were presented to highlight the nature of the design considerations that are important in the development of displays that convey spatial orientation information. Factors emphasized included the need to characterize both optical/visual flow fields and the control dynamics of manual and head-slaved systems.

# REFERENCES

Gibson, J. J.: The perception of the visual world. Houghton Mifflin, 1950.

Johnson, W. W.; Bennett, C. T.; Tsang, P. S.; and Phatak, A. V.: The visual control of simulated altitude. Proc. Fourth Symposium on Aviat. Psychol., 1987, in press.

Koenderink, J. J.; and van Doorn, A. J.: Local structure of movement parallax of the plane. J. Opt. Soc. Amer., vol. 66, 1976, pp. 717-723.

Owen, D. H.; and Warren, R.: Optical variables as measures of performance during simulated flight. Proc. of Human Factors Soc. 27th Ann. Meeting, 1982, pp. 312-315.

Regan, D.; and Beverley, K. I.: Visual flow and direction of locomotion. Science, vol. 227, 1985, pp. 1063-1064.

Uttal, W. R.: The detection of nonplanar surfaces in visual space. Lawrence Erlbaum Associates, Hillsdale, NJ, 1985, pp. 1-22.

Wolpert, L.; Weon, D. H.; and Warren, R.: The isolation of optical information and its metrics for the detection of descent. Optical flow and texture variables useful in simulating self motion (II), Final Technical Report for Grant No. AFOSR-81-0078. Columbus, OH: The Ohio State Univ., Depart. Psychol., Aviation Psychol. Lab., 1983.
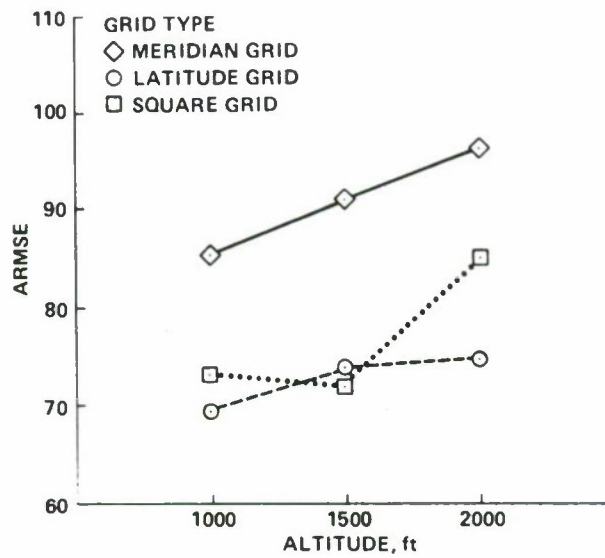
Figure 1.– Mean median ARMSE as a function of grtid type and altitude (ARMSE measures based on 10-sec intervals.
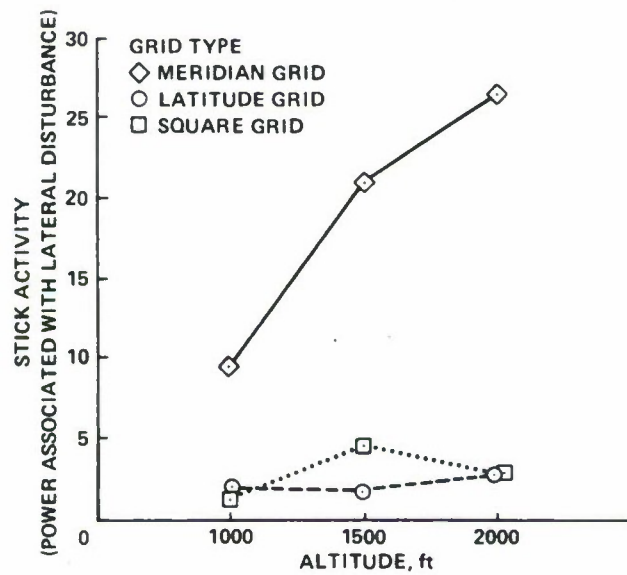


Figure 2.– Control-stick activity associated with lateral disturbance as a function of grid type and altitude (subject 5).
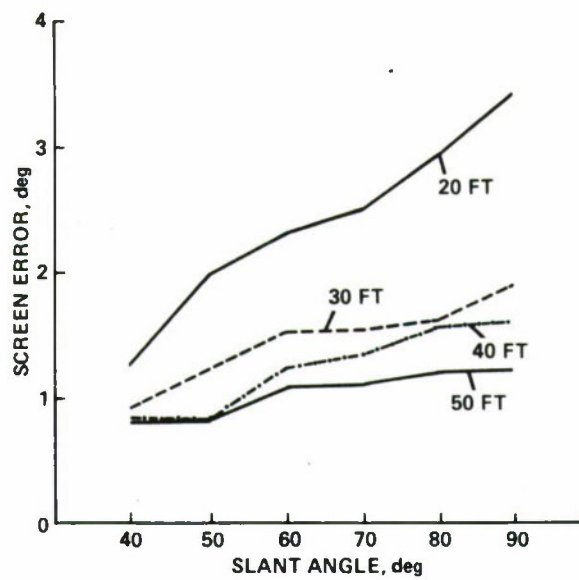
Figure 3.– Head tracking/virtual world, screen error versus offset.
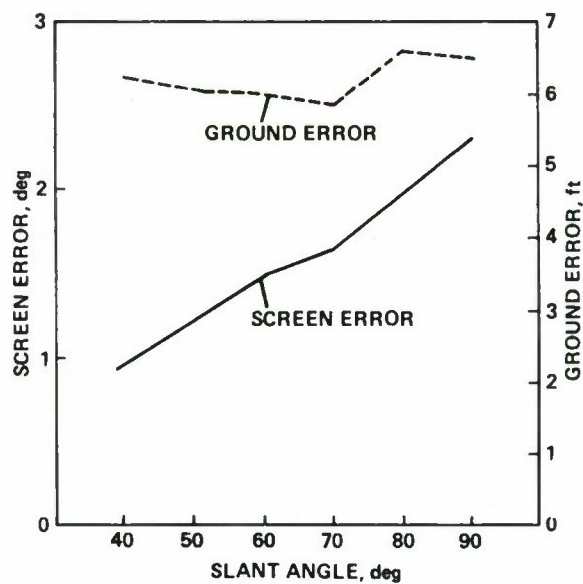


Figure 4.– Head tracking/virtual world (all conditions).
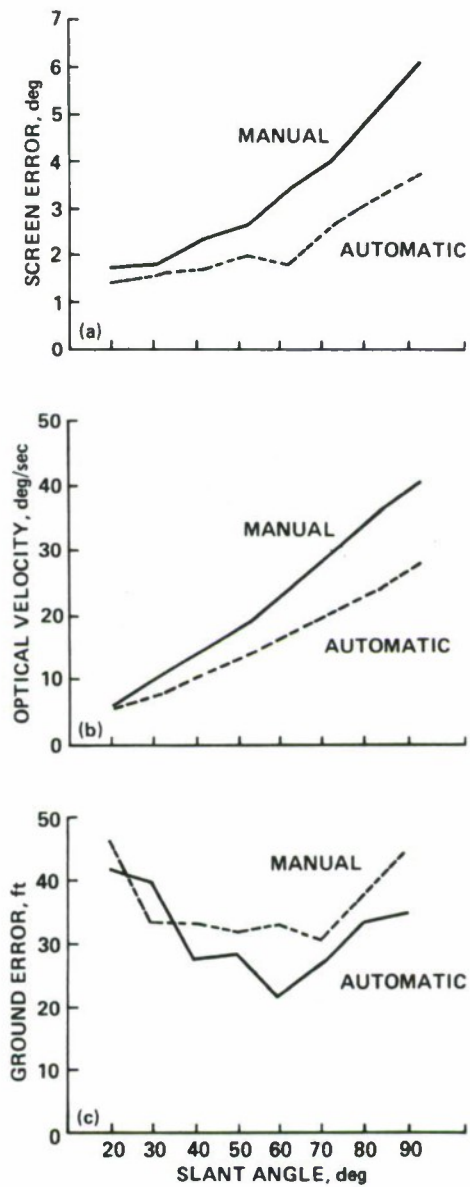
Figure 5.– Manual versus automatic flight, all conditions. (a) Screen error. (b) Optical velocity. (c) Ground error.

# VESTIBULAR ASPECTS

# SENSORY CONFLICT IN MOTION SICKNESS: AN OBSERVER THEORY APPROACH

Charles M. Oman
Man Vehicle Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts

## SUMMARY

"Motion sickness" is the general term describing a group of common nausea syndromes originally attributed to motion-induced cerebral ischemia, stimulation of abdominal organ afferent, or overstimulation of the vestibular organs of the inner ear. Sea-, car-, and airsickness are the most commonly experienced examples. However, the discovery of other variants such as Cinerama-, flight simulator-, spectacle-, and space sickness in which the physical motion of the head and body is normal or absent has led to a succession of "sensory conflict" theories which offer a more comprehensive etiologic perspective. Implicit in the conflict theory is the hypothesis that neural and/or humoral signals originate in regions of the brain subserving spatial orientation, and that these signals somehow traverse to other centers mediating sickness symptoms. Unfortunately, our present understanding of the neurophysiological basis of motion sickness is far from complete. No sensory conflict neuron or process has yet been physiologically identified. To what extent can the existing theory be reconciled with current knowledge of the physiology and pharmacology of nausea and vomiting? This paper reviews the stimuli which cause sickness, synthesizes a contemporary Observer Theory view of the Sensory Conflict hypothesis, and presents a revised model for the dynamic coupling between the putative conflict signals and nausea magnitude estimates. The use of quantitative models for sensory conflict offers a possible new approach to improving the design of visual and motion systems for flight simulators and other "virtual environment" display systems.

## STIMULI CAUSING MOTION SICKNESS: EXOGENOUS MOTION AND "SENSORY REARRANGEMENT"

Motion sickness is a syndrome characterized in humans by signs such as vomiting and retching, pallor, cold sweating, yawning, belching, flatulence, decreased gastric tonus; and by symptoms such as stomach discomfort, nausea, headache, feeling of warmth, and drowsiness. It has a significant incidence in civil and military transportation, and is a common consequence of vestibular disease. Virtually everyone is susceptible to some degree, provided the stimulus is appropriate and lasts long enough. Many other animal species also exhibit susceptibility.

A century ago, physicians commonly attributed motion sickness to acceleration-induced cerebral ischemia, or to mechanical stimulation of abdominal afferents (Reason and Brand, 1975). These theories were largely discounted when the role of the inner ear vestibular organs in body movement control was appreciated, and when James (1882) noted that individuals who lack

vestibular function were apparently immune. As a result, it was commonly thought that motion sickness results simply from vestibular overstimulation.

Certainly the most common physical stimulus for motion sickness is exogenous (i.e., non-volitional) motion, particularly at low frequencies. However, when individuals are able to (motorically) anticipate incoming sensory cues, motion stimuli are relatively benign. For example, drivers of cars and pilots of aircraft are usually not susceptible to motion sickness, even though they experience the same motion as their passengers. In daily life, we all run, jump, and dance. Such endogenous (volitional) motions never make us sick. Thus, it is now recognized that motion sickness cannot not result simply from vestibular overstimulation.

Many forms of motion sickness consistently occur when people are exposed to conditions of "sensory rearrangement"—when the rules which define the normal relationship between body movements and the resulting neural inflow to the central nervous system have been systematically changed (Reason, 1978). Whenever the central nervous system receives sensory information concerning the orientation and movement of the body which is unexpected or unfamiliar in the context of motor intentions and previous sensory-motor experience—and this condition occurs for long enough—motion sickness typically results. Thus, sickness occurs when a person moves about while wearing a new pair of glasses (spectacle sickness) or when a subject in laboratory experiments walks around wearing goggles which cause left-right or up-down reverse vision. Similarly, sickness is also encountered in flight simulators equipped with compelling visual displays (simulator sickness) and in wide-screen movie theaters (Cinerama sickness), since visual cues to motion are not matched by the usual pattern of vestibular and proprioceptive cues to body acceleration. Space sickness among astronauts is believed to result in part because the sensory cues provided by the inner ear otolith organs in weightlessness do not correspond to those experienced on Earth. Astronauts also commonly experience visual spatial reorientation episodes which are provocative. When one floats in an inverted position in the spacecraft, a true ceiling can seem somehow like a floor. Visual cues to static orientation can be ambiguous, often because of symmetries inherent in the visual scene. Cognitive reinterpretation of ambiguous visual orientation cues results in a sudden change in perceived orientation, which astronauts have found can be nauseogenic (Oman, 1988). These various forms of sickness illustrate that the actual stimulus for sickness cannot always be adequately quantified simply by quantifying the physical stimulus. The trigger for sickness is a signal inside the central nervous system (CNS) which also depends on the subject's previous sensory motor experience.

## PHYSIOLOGICAL BASIS OF MOTION SICKNESS

Despite the ubiquity of motion sickness in modern society and significant research (well reviewed, collectively, by Tyler and Bard, 1949; Chinn and Smith, 1955; Money, 1970; Reason and Brand, 1975; Graybiel, 1975; and Miller, 1988), the physiological mechanisms underlying motion sickness remain poorly defined. Classic studies of canine susceptibility to swing sickness (Wang and Chinn, 1956; Bard et al. 1947) indicated that the cerebellar nodulus and uvula—portions of the central vestibular system—are required for susceptibility. Many neurons in the central vestibular system which subserve postural and oculomotor control are now known to respond to a variety of spatial orientation cues, as reviewed by Henn et al. (1980). A brain stem vomiting center was identified by Wang and Borison (1950) and Wang and Chinn (1954), which initiates emesis in dogs in response to various stimuli, including motion. Nausea sensation in humans is

commonly assumed to be associated with activity in the vomiting center (Money, 1970). The integrity of an adjacent chemoreceptive trigger zone (CTZ), localized in area postrema on the floor of the fourth ventricle, was also believed to be required for motion sickness (Wang and Chinn, 1954; Brizzee and Neal, 1954). It was generally assumed that signals originating somewhere in the central vestibular system somehow traverse to the chemoreceptive trigger zone, which in turn activates the vomiting center. Wang and Chinn (1953) and Crampton and Daunton (1983) have found evidence suggestive of a possible humoral agent in cerebrospinal fluid (CSF) transported between the third and fourth ventricle. However, an emetic linkage via CSF transport does not easily account for the very short latency vomiting which is occasionally observed experimentally. The vomiting center receives convergent inputs from a variety of other central and peripheral sources, including the diencephalon and gastrointestinal tract. The possibility of multiple emetic pathways and significant interspecies differences in mechanism must be considered. Also, more recent experiments have led workers to question the notion that medullary emetic centers are discretely localizable. Attempts to verify the earlier findings by demonstrating motion sickness immunity in area postrema ablated and cerebellar nodulectomized and uvulectomized animals have not been successful (Miller and Wilson, 1983a,b; Borison and Borison, 1986; Wilpizeski, Lowry, and Goldman, 1986).

The act of emesis itself involves the somatic musculature. However, many other signs of motion sickness as listed earlier and associated with vasomotor, gastric, and respiratory function suggest that areas in the reticular core of the brain stem and limbic system, which are associated with autonomic regulation are also coactivated. The limbic system and associated hypothalamus-pituitary-adrenal cortex (H-P-A) neuroendocrine outflow pathway is involved. Increases in circulating levels of such stress-related hormones as epinepherine and norepinepherine, ADH, ACTH, cortisol, growth hormone, and prolactin have been found during sickness (e.g., Eversmann et al., 1978; La Rochelle et al., 1982). Whether the limbic system and H-P-A axis simply mediate a generalized stress response, or are also involved in motion-sickness adaptation by somehow triggering stimulus-specific sensory/motor learning is unknown. The question of the site of action of antimotion-sickness drugs is also far from resolved. There is no substantial evidence that effective drugs act on the vestibular end organs. Their primary effect is probably simply to raise the threshold for sickness. Antimotion-sickness drugs could be acting on brain-stem emetic centers. Alternatively, they may shift the fundamental andrenergic-cholinergic balance in the limbic system (e.g., Janowsky et al., 1984).

## DEVELOPMENT OF THE SENSORY CONFLICT THEORY

Although our physiological understanding of motion sickness is thus incomplete, analyses of the wide variety of physical stimuli which produce the same syndrome of symptoms and signs and the dynamic pattern of these responses have nonetheless given us some insight concerning possible etiologic mechanisms. Recognition that motion sickness could occur not only under exogenous motion stimulation, but also as a result of sensory rearrangement, as defined above, has led to the development of a succession of sensory conflict theories for the disorder.

The sensory conflict hypothesis for motion sickness was originally proposed by Claremont (1931), and has since been revised and extended by several authors. Implicit is the idea that a neural or humoral sensory conflict-related signal originates somewhere in the brain and somehow couples to brain centers mediating sickness symptoms. In early statements of the theory, conflict

signals were assumed to somehow result from a direct comparison of signals provided by different sensory modalities (e.g., "the signals from the eye and ear do not agree"; canal-otolith, and visual-inertial conflicts). However, Reason (1978) emphasized that a direct intermodality comparison of afferent signals is simply not appropriate, because signals from the various sense organs have different "normal" behavior (in terms of dynamic response and coding type), and whether they can be said to conflict or not actually depends upon context and previous sensory-motor experience. Hence the conflict is more likely between actual and anticipated sensory signals. Extrapolating from earlier interrelated work by von Holst and Held, Reason argued that the brain probably evaluates incoming sensory signals for consistency using an "efference copy" based scheme. As motor actions are commanded, the brain is postulated to continuously predict the corresponding sensory inputs, based on a neural store (memory bank or dictionary) of paired sensory and motor memory traces learned from previous experience interacting with the physical environment. Sensory conflict signals result from a continuing comparison between actual sensory input and this retrieved sensory memory trace. Any situation which changed the rules relating motor outflow to sensory return (sensory rearrangement, a term coined by Held) would therefore be expected to produce prolonged sensory conflict and result in motion sickness. Adaptation to sensory rearrangement was hypothesized to involve updating of the neural store with new sensory and motor memory-trace pairs. Reason proposed a formal Neural Mismatch model which incorporated these concepts. However, the model was only qualitative, making simulation and quantitative prediction beyond its reach. Key structural elements such as the Neural Store and memory traces were only intuitively defined. The model did not really address the question of why the CNS should have to compute a sensory conflict signal, other than to make one sick. Reason's model dealt with sensory conflict only and did not incorporate emetic brain output pathway elements which must be present to account for the latency and order of appearance of specific symptoms.

## A MATHEMATICAL DEFINITION OF SENSORY CONFLICT

In order to address these difficulties, the author proposed a model for motion sickness (Oman, 1978; 1982) in a mathematical form, shown in block diagram format in figures 1-3. This new model contained a statement of the conflict theory which was congruent with Reason's view, and also the emetic linkage output pathway dynamics missing from Reason's model. The conflict theory portion of the model was formally developed by application of Observer Theory concepts from control engineering to the neural information processing task faced by the CNS in actively controlling body movement using a limited set of noisy sensory signals. The conflict model formulation can be considered an extension of the optimal control model in the field of Manual Control (Baron and Kleinman, 1968) and in the field of spatial orientation research, an extension of Kalman filter models (Young, 1970; Borah, Young, and Curry,1978). The latter have been used to predict orientation perception in passive observers with some success. In these previous models, however, sensory conflict was not defined in the same sense as that used by Reason and me.

In the guidance, control, and navigation systems, engineers are often faced with the problem of controlling a vehicle's state vector (e.g., angular and linear position, velocity, and acceleration) when information from sensors which measure these states is noisy or is even not directly measured at all. To deal with this problem, engineers now routinely incorporate into the control system design a computational element known as an "observer," whose function is to provide an optimal estimate of the actual states of the vehicle (or other system) being controlled. Control loops are closed using the state estimate provided by the observer in lieu of direct feedback sensor

measurements in the traditional way. Analytical techniques have been developed (Kalman, 1960; Wonham, 1968) for mathematically linear systems which allow designers to choose observer and control-loop parameters so that the observer state estimate is always converging with reality, and which optimizes the closed-loop performance of the entire system. In control engineering parlance, such systems are formally called "output feedback" optimal-control systems.

Of particular importance in the present context is the way in which the observer state estimate is calculated in these engineering systems. The observer contains an internal dynamic model of the controlled system and of the sensors being used. The observer element uses these models to calculate what the available feedback sensor measurements should be, assuming the vehicle state estimate of the observer is correct. The difference between the expected and the actual feedback measurements is then computed, because it is an indirect measure of the error in the observer state estimate. The difference signals play an important role in the observer. They are used to continuously steer the observer vehicle state estimate toward reality, using a method described in more detail below.

There is a direct analogy between the "expected" feedback sensor measurement and "internal dynamic model" concepts in control engineering Observer Theory, and the "efference copy" and "neural store" concepts which have emerged in physiology and psychology. From the perspective of control engineering, the "orientation" brain must "know" the natural behavior of the body, i.e., have an internal model of the dynamics of the body, and maintain a continuous estimate of the spatial orientation of all of its parts. Incoming sensory inputs would be evaluated by subtraction of an efference copy signal, and the resulting sensory conflict signal used to maintain a correct spatial orientation estimate.

The mathematical model for sensory conflict and movement control in the orientation brain is shown schematically in figure 2, and mathematically in figure 3. (Arrows in the diagrams represent vector quantities. For example, the actual state of the body might consist of the angular and/or linear displacement of all the parts of the body, and higher derivatives.) The model function can be summarized as follows: the internal CNS models are represented by differential equations describing body and sense organ dynamics. Based on knowledge of current muscle commands, the internal model equations derive an estimated orientation state vector, which is used to determine new muscle commands based on control strategy rules. Simultaneously, the estimated orientation state is used by the CNS sense organ model to compute an efference copy"vector. If the internal models are correct, and there are no exogenous motion disturbances, the efference copy vector nearly cancels polysensory afference. If not, the difference—the sensory-conflict vector—is used to steer the model predictions toward reality, to trigger corrective muscle commands, and to indicate a need for reidentification of the internal model differential equations and steering factors.

How a sensory conflict vector might be used to correct internal model predictions is shown explicitly in figure 3. Here, the physical body and sense organ dynamic characteristics are expressed in linearized state variable notation as a set of matrix equations of the form:

1) $$\dot{\underline{x}} = A\underline{x} + B\underline{u}$$

2) $$\underline{a} = S\underline{x} + \underline{n}_a$$

3) $$\underline{u} = \underline{m} + \underline{n}_e$$

The coefficients of the state differential equations for body and sense organ characteristics are thus embodied in the matrices A, B, and S. These equations are shown graphically in the upper half of figure 3. The internal CNS dynamic model is represented by an analogous state differential equation using hatted variables in the bottom half of the figure. This state estimator (the observer) with its matrices $\hat{A}$, $\hat{B}$, and $\hat{S}$ corresponds to the Neural Store of Reason's (1978) model. The sensory conflict vector $\underline{c}$ is obtained by subtracting actual sensory input $\underline{a}$ from expected sensory input $\hat{S}\,\hat{x}$. Sensory conflict normally originates only from exogenous motion cue inputs $\underline{n}_e$, and noise $\underline{n}_a$. The conflict vector is multiplied by a matrix K calculated using an optimization technique defined by Kalman and Bucy (1961) which lightly weights noisy modalities. When the result is added to the derivative of the estimated state, the estimated state vector is driven toward the actual state, and the component of the conflict vector magnitude due to noise is reduced. However, when exogenous motion cues inputs $\underline{n}_e$ are present, or under conditions of sensory rearrangement, such that matrices A, B, and/or S are changed, and no longer correspond to the matrices of the internal model, actual sensory input $\underline{a}$ will be large, and will not be cancelled by the efference copy vector. Sensory-motor learning takes place via reidentification by analysis of the new relationship between muscle commands and polysensory afference (reidentification of $\hat{A}$, $\hat{B}$, and $\hat{S}$), and internal model updating. Additional details are available in Oman (1982).

This model for sensory conflict overcomes many of the limitations of Reason's Mismatch approach outlined earlier. The Neural Store is replaced by an internal mathematical dynamic model, so that efference copy and sensory conflict signals are quantitatively defined. Increased sensory conflict is noted to result not only from sensory rearrangement, but also from exogenous disturbance forces acting on the body. The role of active movement in creating motion sickness in some circumstances, and in alleviating them in others is clarified.

## A REVISED MODEL FOR SYMPTOM DYNAMICS

The author's 1982 motion-sickness model included dynamic elements in the path between sensory conflict and overall discomfort and nausea in motion sickness. This model has since been altered in some important details; the current version is shown in figures 4 and 5.

The input to the model is a sensory conflict vector. Because of the bandwidth requirements imposed on signals involved in orientation perception and posture control, it seems likely that the components of the conflict vector are neurally coded. In the nausea model, the various conflict vector components (describing the visual, vestibular, proprioceptive modalities) are rectified, and then weighted and added together. Rectification is required because sensory conflict components, as Reason and I have defined them, are signed quantities. The information carried in the sign is

presumably useful in correcting orientation perception and posture control errors. However, stimuli which presumably produce sensory conflicts of opposite signs produce the same type and intensity of nausea, as far as we can tell. Hence rectification is appropriate here. In weighting the various conflict components, vestibular conflicts (i.e., semicircular canal and otolith modalities) must be weighted relatively heavily in the model, since people without vestibular function seem to · be functionally immune. Visual motion inputs (as in Cinerama and simulator sickness) may thus exert their major sick-making effects indirectly: Visual inputs would create illusory movement and thus expected vestibular signals, so sensory conflicts would be produced in the heavily weighted vestibular modality. However, to be consistent with our experimental evidence that visual and proprioceptive conflicts under prism goggle sensory rearrangement (Oman, 1987; Eagon, 1988) eventually become provocative while writing or when building can structures on a desktop, absent concomitant head motion or vestibular conflict, visual and proprioceptive modality model weighting factors are not zero.

As shown in figures 4 and 5, rectified, weighted conflict signals then pass along two parallel, interacting dynamic pathways (fast and slow paths) before reaching a threshold/power law element and resulting in a nausea-magnitude estimate model output. Magnitude estimates are assumed to be governed by a power law relationship (Stevens, 1957) with an exponent of about 2. Susceptibility to motion sickness is determined in the model not only by the amount of sensory conflict produced, but also by the fast and slow pathway gains, time constants, and the nausea threshold. The transfer of a generalized adaptation from one different nauseogenic stimulus situation to another might result from adaptation in these output pathways.

The parallel arrangement of the fast and slow pathways and their relationship to the threshold element requires some explanation. In the past, many authors have therefore assumed that sensory conflict coupling to symptom pathways is a temporary (facultative) phenomenon. However, I have argued (Oman, 1982) that some level of subliminal sensory conflict coupling must be present in normal daily life because conflict signals seem to be continuously functionally averaged at subliminal levels, probably by the same mechanisms or processes which determine the intrinsic dynamics (latency, avalanching tendency, recovery time, etc.) of symptoms and signs when conflict exceeds normal levels. The output pathways probably consist functionally of dynamic elements followed by a threshold, and not the reverse, as would be the case if the linkage were temporary.

In the model, information flows along two paths prior to reaching the threshold. Both paths incorporate dynamic blocks which act to continuously accumulate (i.e., low pass filter or "leaky" integrate) the weighted, rectified conflict signal. One block (the fast path) has a relatively short characteristic response time, and the other (the slow path) has a relatively long one. (In the model simulations shown in the insets of figure 5, the fast path is a second low-pass filter with 1-min time constants; the slow path is a similar filter with 10-min time constants. Second-order or higher block dynamics are required so that model predictions show characteristic overshoot when the conflict stimulus is turned off.) The slow path block normally has a higher gain (by a factor of about 5) than the fast path, and at the beginning of stimulation is functionally the more important element. Slow path output acts together with other classes of fast-acting nauseogenic inputs (e.g., vagal afference from the gut, or emetic drug stimulation) to bias the threshold of nausea response. In the present model, the slow path block output also acts as a multiplicative factor on fast path response gain. When prolonged stimulation has raised the slow path output, the response of the fast path becomes much larger, as shown in the figure 5 simulation. Thus, the revised model mimics the much magnified response to incremental stimulation which we observe experimentally

in long-duration sickness. (In the 1982 version of this model, increased response sensitivity at high symptom levels was a consequence only of the time-invariant, power-law, magnitude-estimation characteristic at the output of the model. This earlier model failed to adequately simulate the rapid rise and fall of sensation at high sickness levels).

Physically, the fast and slow dynamic elements in the model could correspond to physiological mechanisms responsible for conveying conflict-related information from the orientation brain to the emetic brain. Since conflict signals must be rectified, and the dynamics of the fast and slow pathways are qualitatively those of a leaky integration process, it is tempting to think that at least the slow dynamics might involve a humoral mediator and/or a second messenger agent. Alternatively, the dynamics might reflect the action of some diffusion or active transport process, or instead be the intrinsic dynamics exhibited by a network of vomiting center neurons to direct neural or humoral conflict signal stimulation.

## CONCLUSIONS

Over the past decade, the sensory conflict theory for motion sickness has become the generally accepted explanation for motion sickness, because it provides a comprehensive etiologic perspective of the disorder across the variety of its known forms. Motion sickness is now defined as a syndrome of symptoms and signs occurring under conditions of real or apparent motion creating sensory conflict. Symptoms and signs (e.g., nausea, vomiting) are not pathognomonic of the motion sickness syndrome unless conditions of sensory conflict are also judged to be present, since the same symptoms and signs also occur in many other nausea related conditions. Thus, the definition of sensory conflict is implicit in any formal definition of the syndrome. It is essential to define as precisely as possible what is meant by the term sensory conflict. Mathematical models for sensory conflict have sharpened our definitions considerably.

The models presented here capture many of the known characteristics of motion sickness in semi-quantitative fashion. However, they have certain limitations, e.g., the sensory conflict model posits a mathematically linear observer. Although recent experimental data are consistent with the notion that the CNS functions as an observer, there is some evidence that sensory conflict is evaluated in nonlinear ways. Also, the model can only mimic, but not predict, the adaptation process. The model for symptom dynamics does not (yet) incorporate elements which account for observed autogenous waves of nausea at high symptom levels, nor the "dumping" of the fast and slow process pathways when emesis occurs. Models for response pathways mediating other physiologic responses such as pallor, skin temperature, and EGG changes have not yet been attempted.

Do the sensory conflict pathways postulated in the models really exist? Unfortunately, to date no such sensory conflict neuron has been found which satisfies the functional criteria imposed by the current theory. The strongest evidence for the existence of a neural or humoral entity which codes sensory conflict is the ability of the conflict theory to account for and predict the many different known forms of motion sickness. One possibility is that conflict pathways or processes do not exist, but in view of the strong circumstantial evidence, this seems unlikely. There are several alternative explanations:

1. Until recently, there has been surprisingly little discussion of exactly what one meant by the term sensory conflict, so that a physiologist would be able to recognize a "conflict" neuron experimentally. The availability of mathematical models has now changed this situation, and provided a formal definition. However, such models must be presented in ways which physiologists can understand.

2. So far, relatively few animal experiments have been conducted with the specific objective of identifying a conflict neuron. The search has been largely limited to the vestibulo-ocular pathways in the brain stem and cerebellum. Recent evidence suggests that cortex and limbic system are major sites for spatial orientation information processing. Real progress may be limited until orientation research focuses on these areas.

3. Although sensory conflict signals are arguably neurally coded, the conflict linkage mechanisms may have a significant humoral component. If so, a search for the emetic link using classical anatomical or microelectrode techniques will be unsuccessful.

Mathematical characterization of the dynamic characteristics of symptom pathways is a difficult black-box, system-identification problem. The model described above was based only on the character of responses to exogenous motion and sensory rearrangements. Much can potentially be learned from the study of dynamic responses to other classes of emetic inputs, and from studying the influence of behavioral (e.g., biofeedback) and pharmacological therapies.

In other areas of systems physiology and psychology, mathematical models have proven their value by providing a conceptual framework for understanding, for interpreting and interrelating the results of previous experiments, and for planning new ones. Mathematical models can become a useful new tool in motion-sickness research. In the fields of flight simulation and virtual environment displays, simulator sickness is an important practical problem. Models for sensory conflict and motion sickness may become useful tools in the design of these systems.

# REFERENCES

Bard, P., Woolsey, C. W., Snider, R. S., Mountcastle, V. B, and Bromley, R. B. (1947) Delimitation of Central Nervous System Mechanisms Involved in Motion Sickness. Fed Proc 6, 72.

Baron, S., and Kleinman, D. L. (1968) The Human as an Optimal Controller and Information Processor. NASA CR-1151.

Borah, J., Young, L. R., and Curry R. E. (1978) Sensory Mechanism Modelling. USAF ASD Report AFHRL TR 78-83.

Borison H. L., and Borison, R. (1986) Motion Sickness Reflex Arc Bypasses the Area Postrema in Cats. Exp. Neurol. 92:723.

Brizzee, K. R., and Neal, L. M. (1954) A Reevaluation of the Cellular Morphology of the Area Postrema in View of Recent Evidence of a Chemoreceptive Function. J. Comp. Neurol. 100, 41.

Chinn, H. I., and Smith, P. K. (1955) Motion Sickness. Pharm. Rev. 7, 33-82.

Claremont, C. A. (1931) The Psychology of Seasickness. Psyche 11, 86-90.

Crampton, G. H., and Daunton, N. G. (1983) Evidence for a Motion Sickness Agent in Cerebrospinal Fluid. Brain Behav. Evol. 23:36-41.

Eagon, J. C. (1988) Quantitative Frequency Analysis of the Electrogastrogram During Prolonged Motion Sickness. M.D. Thesis, Harvard-MIT Div. of Health Sci. and Technol., Massachusetts Institute of Technology, Cambridge.

Eversmann, T., Gottsmann, M., Uhlich, E., Ulbrecht, G., von Werder, K., and Scriba, P. C., (1978) Increased Secretion of Growth Hormone, Prolactin, Antidiuretic Hormone, and Cortisol Induced by the Stress of Motion Sickness. Aviat. Space Environ. Med. 49:53-57.

Graybiel, A. (1975) "Angular Velocities, Angular Accelerations, and Coriolis Accelerations." Ch. 7 in Foundations of Space Biology and Medicine, Vol. II, Book 1, NASA/USSR Acad. Sci., Washington, DC.

Janowsky, D. S., Risch, S. C., Ziegler, M., Kennedy, B., and Huey, L. (1984) A Cholinomimetic Model of Motion Sickness and Space Adaptation Syndrome. Aviat. Space Env. Med. 55:692-696.

James, W. (1882) The Sense of Dizziness in Deaf Mutes. Amer. J. Otol. 4, 239-25.

Kalman, R. E. (1960) Contributions to the Theory of Optimal Control. Biol. Soc. Mat. Mexicana 5, 102-119.

Kalman, R. E., and Bucy, R. S. (1961) New Results in Linear Filtering and Prediction Theory. J. Basic. Eng. Trans. ASME, Ser. D 83, 95-108.

La Rochelle, F. T. Jr., Leach, C. S., Homick, J. L., and Kohl, R. L. (1982) Endocrine Changes During Motion Sickness: Effects of Drug Therapy. Proc. 1982 Aerosp. Med. Assn. Meeting.

Miller, A. D. (1988) "Motion-Induced Nausea and Vomiting." Ch. 5 in: Nausea and Vomiting: Recent Research and Clinical Advances. Harding, R. K., Kucharczyk, J., and Stewart, D. J. (eds.), CRC Press.

Miller, A. D., and Wilson V. J. (1983a) Vestibular-Induced Vomiting After Vestibulo-Cerebellar Lesions. In: Mechanisms of Motion Induced Vomiting. Brain Behav. Evol. 23, 26-31.

Miller, A. D., and Wilson, V. J. (1983b) "Vomiting Center" Reanalyzed: An Electrical Stimulation Study. Brain Res. 270, 154-158.

Money, K. E. (1970) Motion Sickness. Physiol. Rev. 50:1-39.

Oman, C. M. (1978) A Sensory Motor Conflict Model for Motion Sickness. Workshop III. Space Motion Sickness Symposium, Nov. 16, 1978, NASA Johnson Space Center, Houston, TX.

Oman, C. M. (1982) A Heuristic Mathematical Model for the Dynamics of Sensory Conflict and Motion Sickness. Acta Otolaryng. Suppl. 392.

Oman, C. M. (1988) "The Role of Static Visual Cues in the Etiology of Space Motion Sickness." In: Igarashi, M. and Nute, K. G. (eds.), Proc. Symposium on Vestibular Organs and Altered Force Environment, Oct. 1987. NASA/Space Biomedical Research Institute, USRA/Division of Space Biomedicine, Houston, TX, pp. 25-38.

Oman, C. M. (1987) Spacelab Experiments on Space Motion Sickness. Acta Astronaut. 15:55-66.

Reason, J. T. (1978) Motion Sickness Adaptation: A Neural Mismatch Model. J. Roy. Soc. Med. 71, 819-829.

Reason, J. T., and Brand, J. J. (1975) Motion Sickness. Academic Press, London.

Stevens, S. S. (1957) On the Psychophysical Law. Psychol. Rev. 64:153-184.

Tyler, D. B., and Bard, P. (1949) Motion Sickness. Physiol. Rev. 29, 311-369.

Wang, S. C., and Borison, H. L. (1950) The Vomiting Center: A Critical Experimental Analysis. Arch. Neurol. Psychiat. 63, 928-941.

Wang, S. C., and Chinn, H. I. (1953) Vestibular Reflex Pathway in Experimental Motion Sickness in Dogs. Abstr. XIX Int. Physiological Congr. (Montreal), pp. 868-869.

Wang, S. C., and Chinn, H. I. (1954) Experimental Motion Sickness in Dogs: Functional Importance of Chemoreceptive Emetic Trigger Zone. Am. J. Physiol. 178, 111-116.

Wang, S. C., and Chinn, H. I. (1956) Experimental Motion Sickness in Dogs: Importance of Labyrinth and Vestibular Cerebellum. Am. J. Physiol. 185, 617-623.

Wilpizeski, C. R., Lowry, L. D., and Goldman, W. S. (1986) Motion-Induced Sickness Following Bilateral Ablation of Area Postrema in Squirrel Monkeys. Laryngoscope 96:122.

Wonham, W. H. (1968) On the Separation Theorem of Stochastic Control. SIAM J. Cont. 6, 312-326.

Young, L. R. (1970) On Visual Vestibular Interaction. Proc. 5th Symp. on Role of the Vestibular Organs in Space Exploration, NASA SP 314.
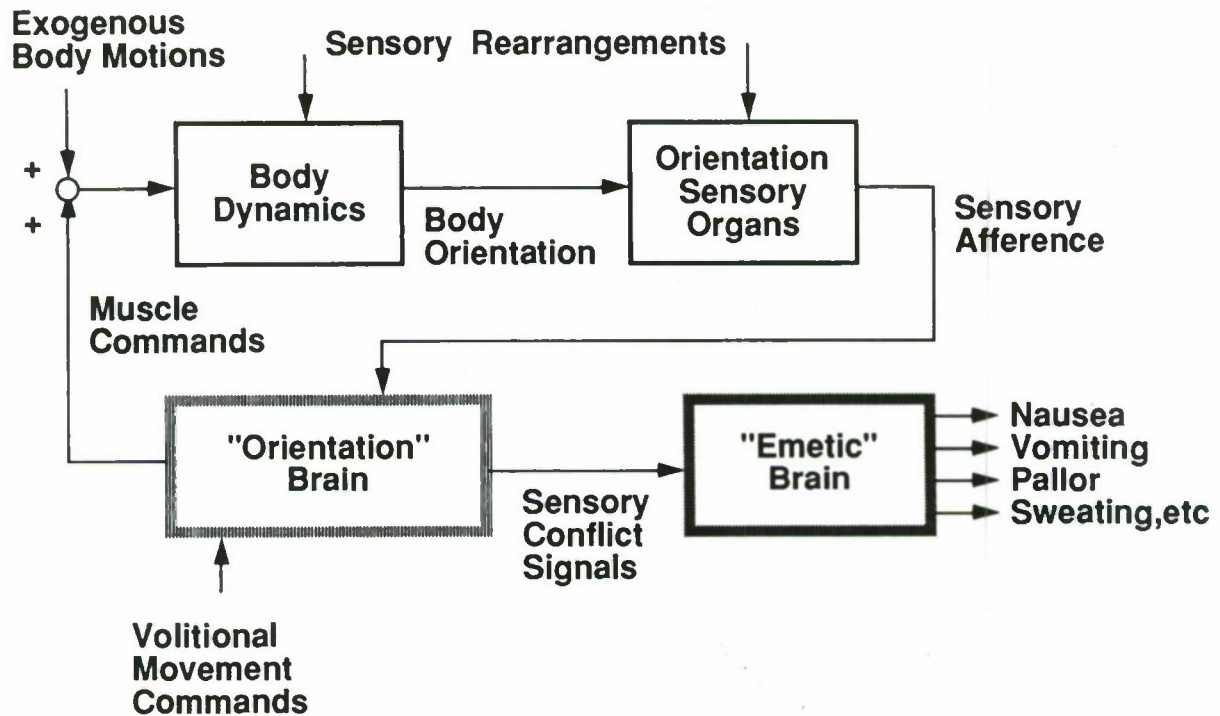
Figure 1.– Schematic diagram of model for movement control, sensory conflict, and motion-sickness symptom dynamics (Oman, 1982). Under conditions of sensory rearrangement, the rules which relate muscle commands to sensory afference are systematically changed. Sensory conflict signals used spatial orientation perception and movement control in the orientation brain couple to the emetic brain.
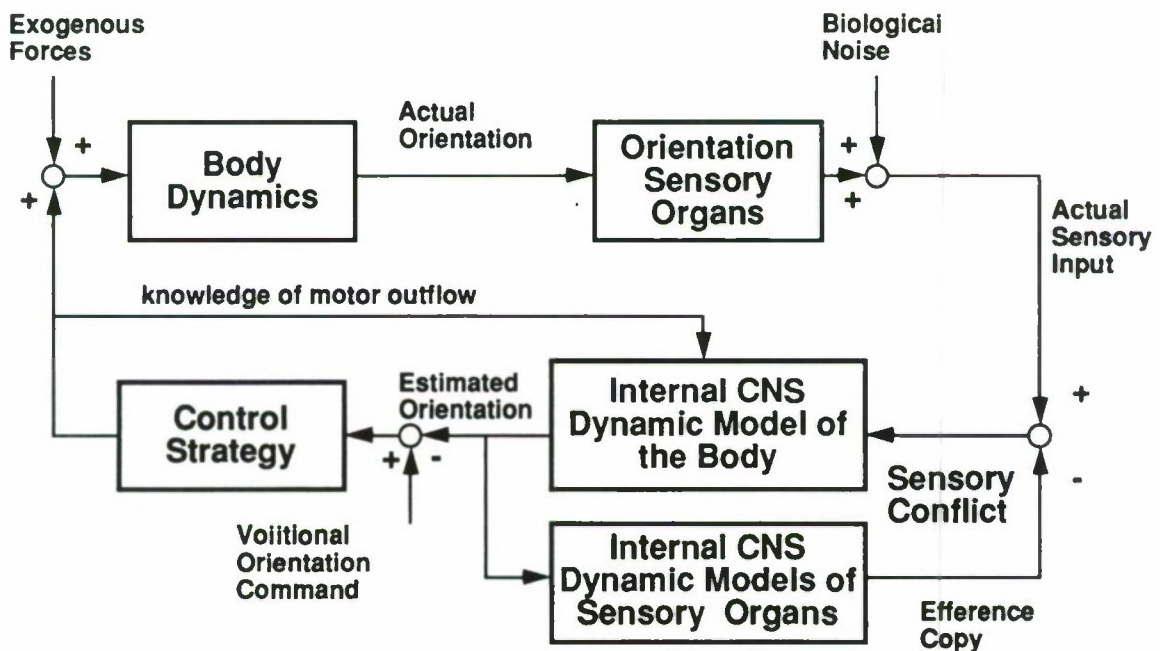


Figure 2.– Observer theory model for movement control (Oman, 1982).
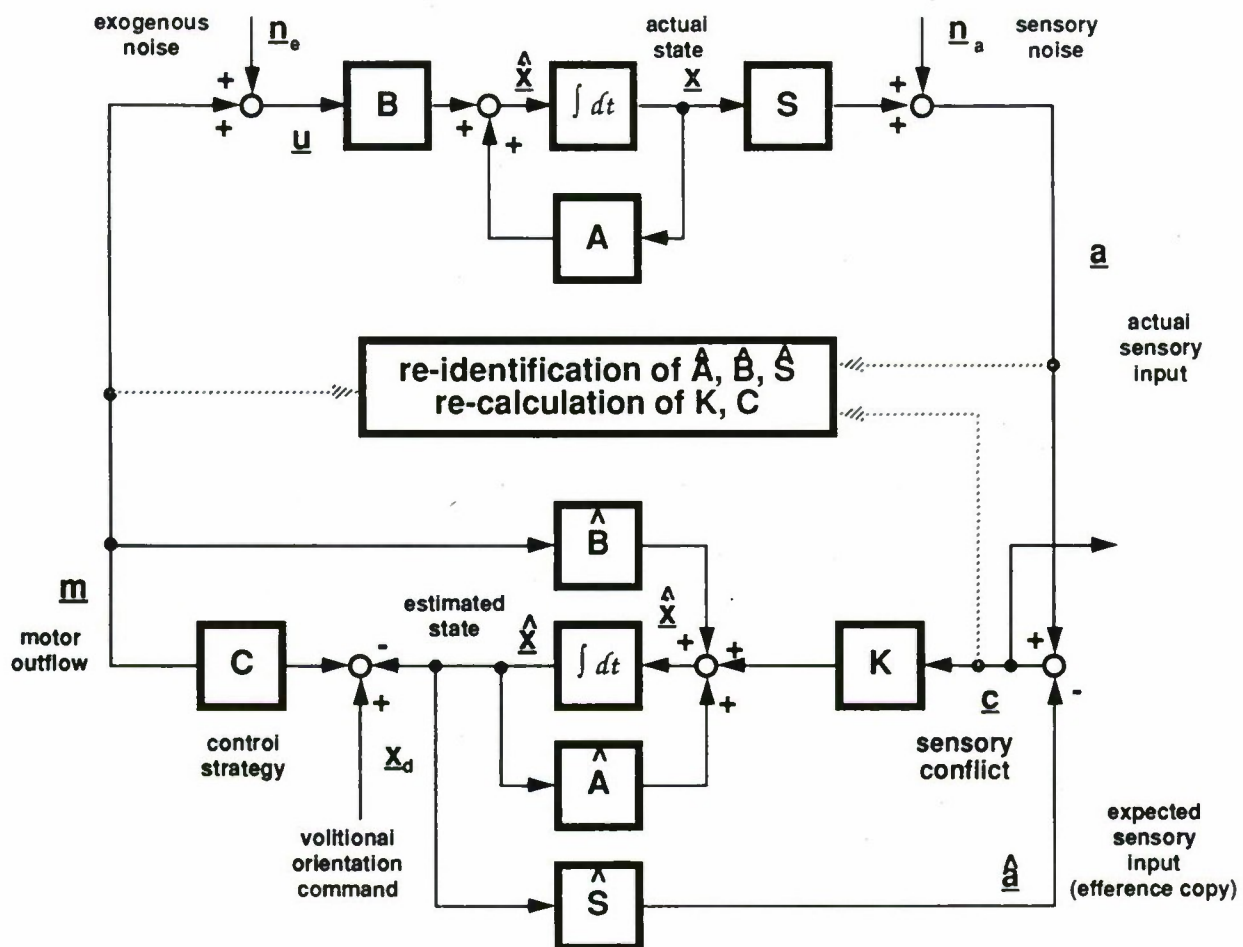
Figure 3.– Mathematical formulation of model shown in figure 2 (Oman, 1982).

**Fast path:**

■ At high nausea levels, a single conflict stimulus produces a virtually instantaneous increment in nausea.

■ therefore likely neurally mediated.
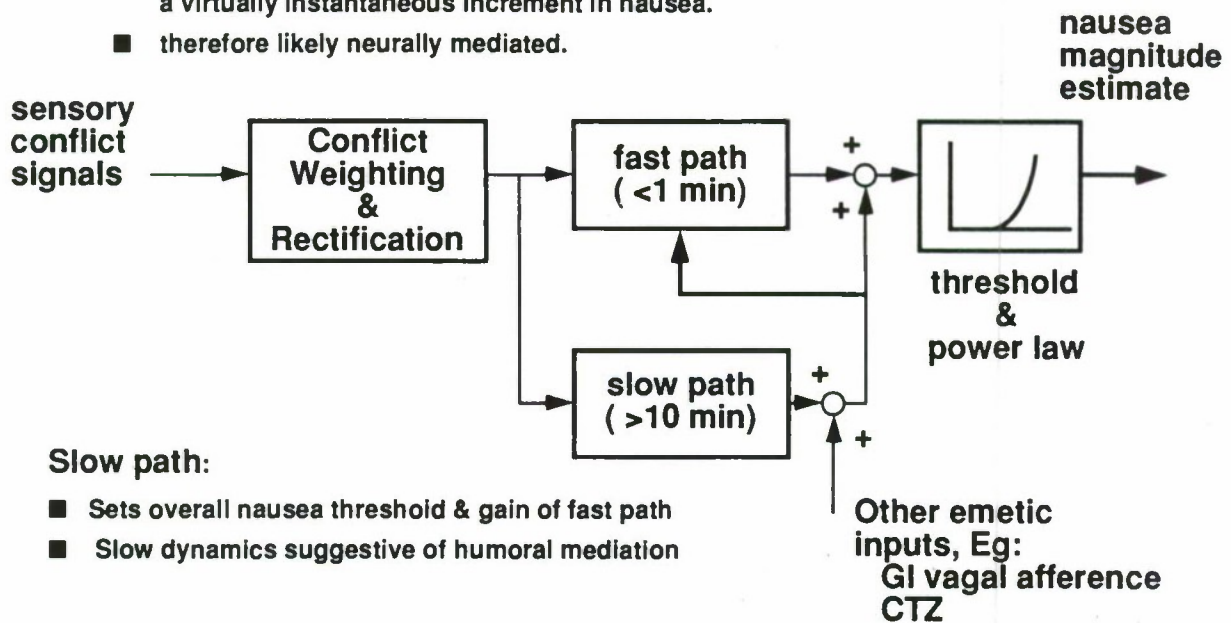
**Slow path:**

■ Sets overall nausea threshold & gain of fast path

■ Slow dynamics suggestive of humoral mediation

Figure 4.– Schematic diagram of revised model for nausea-path symptom dynamics.



**Stimulus:**

**3 successive 10 min intervals at a "moderate" level**

Eg: head movements (5 RPM Coriolis or with prism goggles)
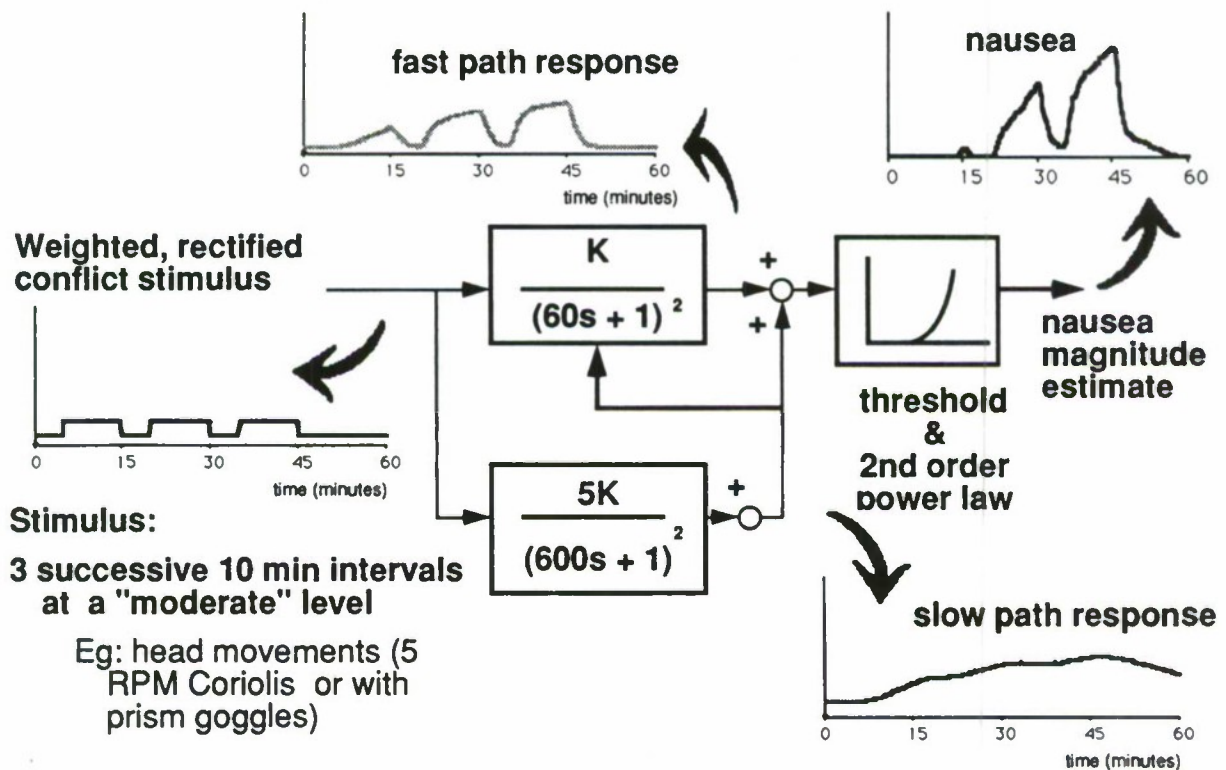
Figure 5.– Mathematical model for nausea-path symptom dynamics. Insets show results of computer simulation.

# INTERACTIONS OF FORM AND ORIENTATION

Horst Mittelstaedt
Max-Planck-Institut fur Verhaltensphysiologie, D-8130 Seewiesen
Bundes Republik Deutschland

## 1. EFFECT OF ORIENTATION OF PERCEPTION OF FORM

It is well know that the orientation of an optical pattern relative to egocentric or extraneous references affects its figural quality, that is, alters its perceived form and concomitantly delays or quickens its identification (Rock 1973). A square presented in the frontal plane to an upright person (S), for instance, changes from a "box" to a "diamond" when it is rotated with respect to the S's median plane by 45°. This angle, that is, the angle between the orientations of the pattern in which the two apparent figures ("Gestalten") attain a summit of purity and distinctness, will be called the "figural disparity" of the pattern. If, as in this case, the S is upright, the retinal meridian and the subjective vertical (SV) are both in the viewer's median plane. The question arises with respect to which of these orientation references the two figures are identified. The answer may be found when the pattern and the S are oriented in such a way that the projections of the retinal meridian and the SV into the plane of the pattern diverge by the pattern's figural disparity or its periodic multiples; that is, in the case of a square by 45° or 135°, respectively. Similarly, which reference determines whether an equilateral triangle is seen as a "pyramid" or a "traffic warning sign" may be revealed at a divergence of SV and retinal meridian of 60° or 180°, respectively. It is generally found that for head roll tilts (ρ) and figural disparities of up to 90°, the figure whose axis coincides with the SV is seen. At head tilts of ρ = 180°, however, the retinal reference dominates, as a rule independently of the figural disparity (for reviews, see Rock 1973 and Howard 1982).

## 2. EFFECT OF FORM ON PERCEPTION OF ORIENTATION

Clearly, then, orientation may determine apparent form. But conversely, form may also influence apparent orientation. This is explicitly true in the case of the SV (for review, see Bischof 1974; for the recent state, see Wenderoth 1976; Mittelstaedt 1986).

As shown in Fig. 1, our method is to project the pattern within a circular frame (of 16°, 35°, or 80° visual angle) into a tilted planetarium cupola (ø = 9.1 m) in 24 stationary orientations presented to the S in a pseudo-random sequence. The S, lying on her side, indicates her SV by means of a rotatable luminous line, which is projected onto the cupola such that its center of rotation coincides with the center of the pattern's circular frame and the S's visual axis.

The effect of the pattern on the SV turns out to be a rather involved function of the orientation of the pattern. This relation becomes clear, however, if we assume that the luminous line is eventually oriented such that the effect of the pattern is opposite and equal to the nonvisual effect on the SV, exerted mainly by the vestibular system. Both effects are then expected to be functions of the difference between the angle $\beta$ at which the luminous line is set with the pattern present and the angle $\beta_g$ at which it is found in the absence of visual cues. For the nonvisual effect, fortunately, this function may be computed according to an extant theory (Mittelstaedt 1983a,b): the SV is influenced not only by information about head tilt, but also by intrinsic parameters which are independent of head tilt, notably the "idiotropic vector" (M). Presumably by addition of constant endogeneous discharges to the saccular output, it leads to a perpetual shift of the SV into the

direction of the S's long axis and hence causes the phenomenon which is well known as the Aubert phenomenon. At first approximation, this relation may be represented by a vector diagram (Fig. 2): In the absence of visual cues, the SV is perceived in the direction of the resultant $R$ of the otolithic vector $G$ and the idiotropic vector $M$.

In our case, since $\rho = 90°$, the nonvisual effect $g$ becomes a particularly simple function $\beta - \beta_g$, namely,

$$g = \sqrt{G^2 + M^2} \sin (\beta - \beta_g) = \sqrt{G^2 + M^2} \sin \left(\beta - \text{arccotan} \frac{M}{G}\right)$$

$$= M \sin \beta - G \cos \beta$$

(1)

Because of the normalization of the vestibular information (which is inferred from effects of centrifugation), $g$ may be computed with $G = 1$ and $M = \text{cotan } \beta_g$. Hence the unknown visual effect on the SV may be determined if the known quantity $g$ is plotted as a function of the angle on which effect of the pattern depends. There seem to be only two possible candidates: the angle $\vartheta$ between the pattern's main axis and the S's long axis, or the angle $\beta - \vartheta$ between the former and the present direction $\beta$ of the SV.

Figure 3 shows plots of this latter function (named SV-function) engendered in three Ss by a color slide of the house of Fig. 1. It turns out that the visual effect is zero, that is, does not change the SV ($\beta = \beta_g$) if and only if $\beta - \vartheta$ is zero, rather than when $\vartheta$ is zero. Hence its magnitude must be a function of the former angle. We may envisage the SV as being at equilibrium between two tendencies ("torques"), (1) the gravito-idiotropic torque $g$, trying to pull it toward $\beta - \beta_g = 0$, (2) the other, the visual torque $\upsilon$, trying to pull it toward $\beta - \vartheta = 0$ (see Fig. 2). Generally, the visual torque exerted on the SV by a pattern turns out to be an antisymmetrical periodic function composed of the sine of $(\beta - \vartheta)$ and the sine of the angle's multiples. Hence it may be simply and fully characterized by the amplitudes $V_n$ of these sine components, to be called "(circular) harmonics" of the respective SV function. With the picture of the house of Fig. 1 as well as with other photographed scenes, the *first* circular harmonic is generally found to vary greatly inter- as well as intrapersonnally. By contrast, the second and fourth harmonics vary but moderately (within an order of magnitude) between Ss, and are rather constant intrapersonnally for a given pattern.[1] The formal difference is supposed to be due to a difference in the underlying information processing. The first harmonic expresses the effect of the picture's bottom-to-top polarity, that is, of those cues for the vertical which may be inferred from its normal orientation to gravity. The recognition of what is the top must probably be learned through personal experience, and its effect is hence expected to vary with individual visual proficiency. The even-number harmonics, by contrast, are presumably based on invariant structures of the visual system, possibly by a weighting process, from the "simple cells" of the visual cortex (Mittelstaedt 1986).

This is highlighted by the following experimental series. If orthogonal lines are presented as a pattern, the resulting SV-function contains only circular harmonics which are multiples of four.

---

[1]All circular harmonics higher than the fourth, except for the eighth, which is sometimes found to be just above noise level, are insignificant or zero. With the sampling used, the amplitudes of the first four harmonics were about the same irrespective of whether the Fourier analysis was made with the equidistant sampling of plots over $\vartheta$ or with the, necessarily, scattered sampling of plots over $\beta - \vartheta$ as in Fig. 3.

The fourth usually is then the largest and is positive; that is, at its null-crossings with positive slope the SV coincides (is in phase) with the direction of the lines (Fig. 4).

If a pictograph of a human figure is presented which consists of uniformly oriented lines (Fig. 5; "star man") or random dashes, the first harmonic is in phase with star man's long axis and hence is positive.

What will happen if the pictograph of a human figure is presented which consists, as in Fig. 6 ("diamond man"), exclusively of lines that are oriented at 45° with respect to the figure's long axis? As a matter of fact, the two figural components are superimposed: the first harmonic is in phase and hence positive; the fourth is in counterphase and hence negative, neither "taking notice" of the other (Fig. 7).

Evidently, the result falsifies the hypothesis (Bischof and Scheerer 1970) that the CNS first computes a "resultant visual vertical" of the picture and subsequently forms an antisymmetrical periodic function *in phase* with this resultant. For then, the resultant would either coincide with the long axis of diamond man and hence the fourth harmonic would be positive, or (rather unlikely though) the resultant would coincide with one of the line directions and hence the first harmonic would be in phase with that line (or would be missing). Instead, the first harmonic results from a processing which is determined by the bottom-to-top polarity of the picture independently of its unpolarized axial features. At the same time, the even-number harmonics are determined by the pattern's unpolarized axial features independently, at least with respect to phase, of its bottom-to-top polarity.

## 3. INTERRELATIONS BETWEEN THE DETERMINANTS OF APPARENT VERTICAL AND OF FORM PERCEPTION

It shall now be examined whether, by means of the comprehensive mathematical theory of the SV, understanding the effect of perceived form on the SV may help in understanding the effect of the SV on form perception mentioned earlier.

First, the theory does indeed offer a good reason why the influence of the SV on the perception of form should decrease with an increasing tilt angle of the S. The effect of the otolithic output than decreases (besides due to comparatively small deviations from a linear response to shear) as a consequence of the addition of the idiotropic vector. Its amount is an idiosyncratic constant averaging around 50% of that of G. The magnitude of the resultant R of the idiotropic vector M and the gravity vector G may be approximated as

$$R = \sqrt{G^2 + M^2 + 2\,GM \cos \rho} \qquad (2)$$

Evidently, R must decrease with increasing angle of tilt ρ, and so will its relative influence when competing with visual cues!

Second, the theory may open a way to assess the relative strength of the factors that influence form perception. The influence of visual patterns on the SV is not independent of the angle of tilt (Bischof and Scheerer 1970). This effect may be quantitatively described by weighting the visual torque v with the sum of the squared saccular and utricular (roll) components (for details see

Mittelstaedt 1986). Hence the effect of the visual torque is maximal at a roll tilt between 60 and 90° and declines toward the upright as well as toward the inverted posture. As a result, at small roll tilts of the S, the nonvisual torque g may, under certain conditions, be larger than the visual torque, about equal to the latter around $\rho = 90°$, but much smaller than the visual torque when the S is inverted ($\rho = 180°$). Which component will determine which form is perceived under which angle of divergence may be predictable, if the relative weights of the nonvisual and visual components in the determination of the SV would be correlated with the relative weights of the two reference systems in the perception of form.

## 4. SUPPRESSION OR ADDITIVE SUPERPOSITION

However, the underlying information-processing systems may be fundamentally different in the two cases. Evidently, additive superposition suffices to explain the interaction of the components in the case of the SV. But in their influence on form perception, a decision in case of conflict appears to be called for, and hence to necessitate a nonlinear interaction in that one of the competitors is suppressed.

This we have tested by using the well-known ambiguous figure of Fig. 8. It is seen, by an upright S, as a "princess" P or a "witch" W when the long axis of P is aligned or reversed with respect to the S's long axis.

If the S is tilted by 180° relative to gravity ($\rho = 180°$) the *retinal* reference determines the perception, as is generally found in comparable cases. The crucial situation arises when the S views the figure while lying on the ($\rho = 90°$). In this position the figure was presented at various angles $\upsilon$ with respect to the S's long axis, and the S was instructed to report whether the witch or the princess appeared more distinctly. In order to determine the point of transition between the two phenomena, their distinctness was scaled by the Ss in seven steps, which are condensed in Fig. 9 into five (exclusively P; preponderantly P; ambiguous; preponderantly W; exclusively W).

Two Ss, who were well versed in psychophysical tests were chosen. In addition their SV in the absence of visual cues and their ocular counterroll at $\rho = 90°$ were determined and were found as shown in Fig. 9. Clearly, in both Ss, the midline between the transition zones neither coincides with the SV nor with the retinal meridian, but assumes an intermediate direction between these two. Hence even in their influence on form perception the gravito-idiotropic and the visual effects may combine vectorially rather than suppress one another.

It is advisable, then, to reexamine those instances where an exclusive decision between the two references is found. As mentioned earlier, this happens regularly, when S and pattern are placed such that *the SV and the retinal meridian diverge by the figural disparity angle.* Now let the "salience" s (die "PRAEGNANZ") of a figure (X) vary as a symmetrical periodic function of its deviation from the respective reference such that

$$s_X = \sum_{n=0}^{max} E_{xn} \cos n\,\vartheta_X' + \sum_{n=0}^{max} V_{xn} \cos n(\beta - \vartheta_X') \tag{3}$$

where $\vartheta$ is the angle between the figure's main axis and the retinal meridian, $\beta'$ is the angle between the SV and the retinal meridian, and $E_{xn'}$, $V_{xn}$ are the amplitudes of the figure's circular

harmonics weighted (as suggested in section 3) by the retinal ($E_{xn}$) and the SV reference systems ($V_{xn}$), respectively. The central nervous correlate of the relative salience of figures X,Y may then be determined by the difference $s_x - s_y$. In the case of princess versus witch, because $\vartheta'_w = \vartheta'_p - 180°$ and if, for the sake of simplicity, nmax is assumed to be unity, the difference becomes

$$s_p - s_w = (E_{p0} + V_{p0}) - (E_{w0} + V_{w0}) + (E_{p1} + E_{w1}) \cos \vartheta'_p + (V_{p1} + V_{w1}) \cos (\beta' - \vartheta'_p) \quad (4)$$

In the upright S ($\rho = 0$, $\beta' = 0$), with $E_{p0} + V_{p0} = E_{w0} + V_{w0}$, this becomes

$$s_p - s_w = (E_{p1} + E_{w1} + V_{p1} + V_{w1}) \cos \vartheta'_p \quad (5)$$

That is, independently of the relative weights, the princess dominates at acute angle $\vartheta'_p$ and the witch dominants at obtuse angles $\vartheta'_p$ However, with the S inverted ($\rho = 180°$, $\beta' = 180°$):

$$s_p - s_w = [(E_{p1} + E_{w1}) - (V_{p1} + V_{w1})] \cos \vartheta'_p \quad (6)$$

Consequently, the pattern is identified exclusively according to one of the two reference systems, if their respective weighting factors differ and $\vartheta \neq 90°$, *even though the assumed processing is purely additive*. The same holds for the other examples given above. In the case of the square, for instance, with n = 4, and the S tilted until $\beta' = 45°$ ($\rho = 45°$),

$$s_{b(box)} - s_{d(diamond)} = [(E_{b4} + E_{d4}) - (V_{b4} + V_{d4})] \cos 4\vartheta'_b \quad (7)$$

This leads to a "decision" in favor of the SV-reference if—quite plausibly at that acute angle—the V factors are then larger than the E factors, whereas at $\beta' = 135°$ ($135° < \rho < 180°$) they appear to be almost equal: in that position some of our Ss refused to decide about what they see! In the case of princess versus witch with the S at $\rho = 90°$ and $\beta = 60°$

$$s_p - s_w = (E_{p1} + E_{w1})\cos \vartheta'_p + (V_{p1} + V_{w1}) \cos (60° - \vartheta'_p) \quad (8)$$

Hence a compromise is to be expected depending on the relative magnitudes of the weighting factors. The relative salience ($s_p - s_w$) is then zero at $\vartheta'_{\Delta p \; zero}$, and

$$\cotan\vartheta'_{\Delta p \; zero} = \frac{\pm\sin 60°}{+\left(\dfrac{E_{p1} + E_{w1}}{V_{p1} + V_{w1}} + \cos 60°\right)} \quad (9)$$

as is borne out by the results here and in Fig. 9.

In conclusion, the present state favors the notion that angular relations are represented and processed in the CNS by variables which are trigonometric functions of the respective angles. That the characteristics and the spatial arrangement of the otolithic receptors and of the simple cells in the visual cortex are well suited to implement this kind of coding (Mittelstaedt 1983a,b; 1986; 1988 in

press) lends a neurophysiological backbone to the demonstrated descriptive and predictive powers of such a theory.

## Acknowledgments

# REFERENCES

Bischof, N. (1974): Optic-vestibular orientation to the vertical. In: Handbook of sensory physiology (Autrum, H. et al., eds.) 155-190 Springer, Berlin, Heidelberg, New York.

Bischof, N. and Scheerer, E. (1970): Systemanalyse der optisch-vestibularen Interaktion bei der Wahrnehmung der Vertikalen. Psychologische Forschung 34, 99-181.

Howard, I. P. (1982): Human visual orientation. Wiley & Sons, New York, Brisbane, Toronto.

Mittelstaedt, H. (1983a): A new solution to the problem of the subjective vertical. Naturwissenschaften 70, 272-281.

Mittelstaedt, H. (1983b): Towards understanding the flow of information between objective and subjective space. In: Neuroethology and behavioral physiology (Huber, F. and Markl, H., eds.) 382-402, Springer, Berlin, Heidelberg, New York.

Mittelstaedt, H. (1986): The subjective vertical as a function of visual and extraretinal cues. Acta Psychologica 63, 63-85.

Mittelstaedt, H. (1988 in press): The information processing structure of the subjective vertical. A cybernetic bridge between its psychophysics and its neurobiology. In: Processing structures for perception and action (Marko, H., Hauske, G., Struppler, A., eds.). Final report of the Sonderforschungsbereich "Kybernetik." Verlag Chemie, Weinheim.

Rock, I. (1973): Orientation and Form. New York: Academic Press.

Wenderoth, P. M. (1976): An analysis of the rod-and-frame illusion and its variants. In: Studies in Perception (Day, R. H., Stanley, G. S., eds.) Perth: Univ. Western Australia, Perth.
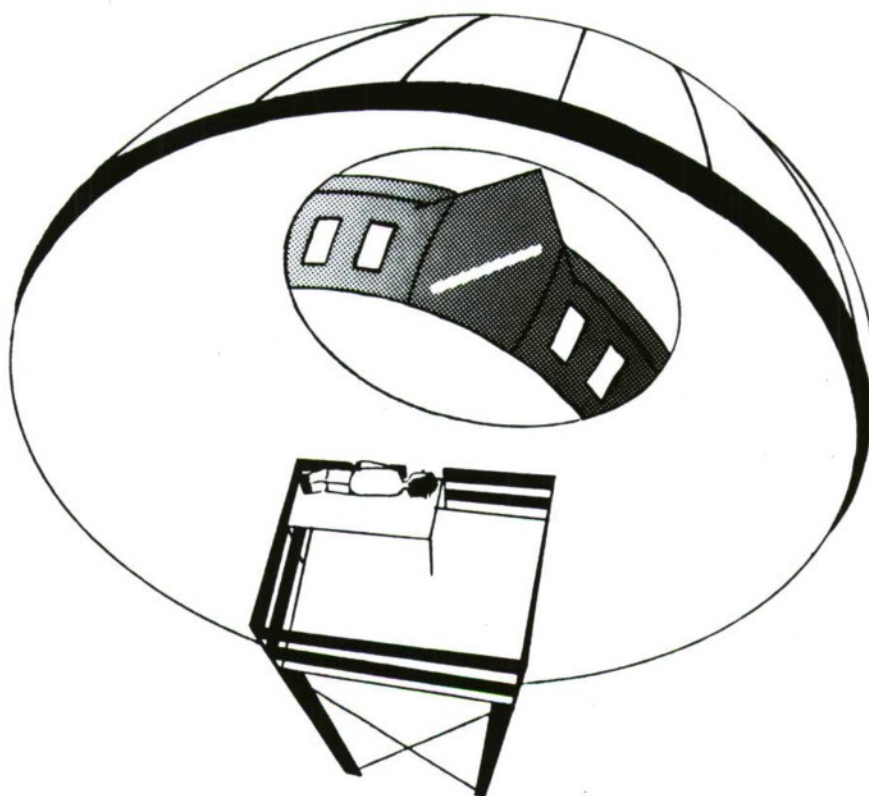
Figure 1.– Experimental setup for testing the effect of tilted images on the subjective vertical. The image is projected in a sequence of static roll tilts onto a hemispherical (ø = 9.1 m) screen in front of the subject. The S, lying on her side, is asked to set a projected luminous line to subjective vertical.
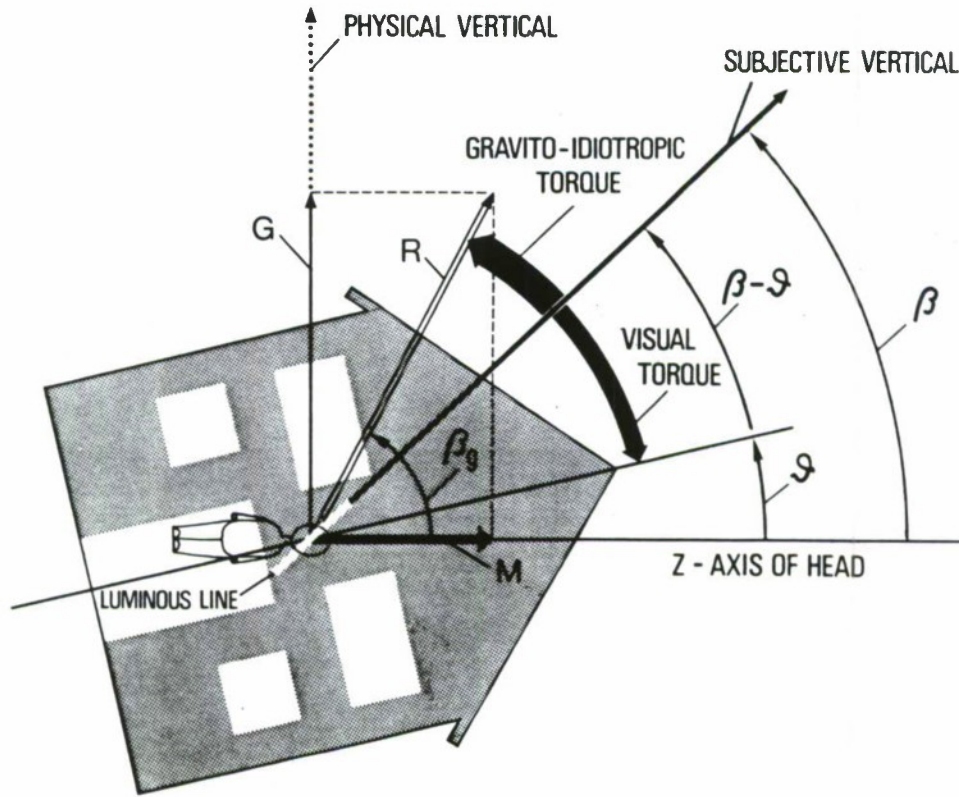
Figure 2.– Definition of critical variables and their relations to hypothetical determinants of the SV: 1) It is supposed that the visual scene (here a house) exerts an attraction effect on the SV. This "visual torque" is supposed to be a function of $\beta - \vartheta$, the angle between the main axis of the tilted image and the luminous line when set subjectively vertical; 2) This visual torque is supposedly counterbalanced by a "gravito-idiotropic torque." The latter is a function of $\beta - \beta_g$, the angle between the present SV and the $\beta_g$ the SV would have in the absence of visual cues. The latter function may be determined as

$$g = \sqrt{G^2 + M^2}\, \sin(\beta - \beta_g) = \sqrt{G^2 + M^2}\, \sin\!\left(\beta - \operatorname{arccotan} \frac{M}{G}\right) = M\, \sin\beta - G\, \cos\beta$$

with $G = 1$. Hence the unknown visual torque may be quantitatively described. All angles defined with respect to (long) Z-axis of head.
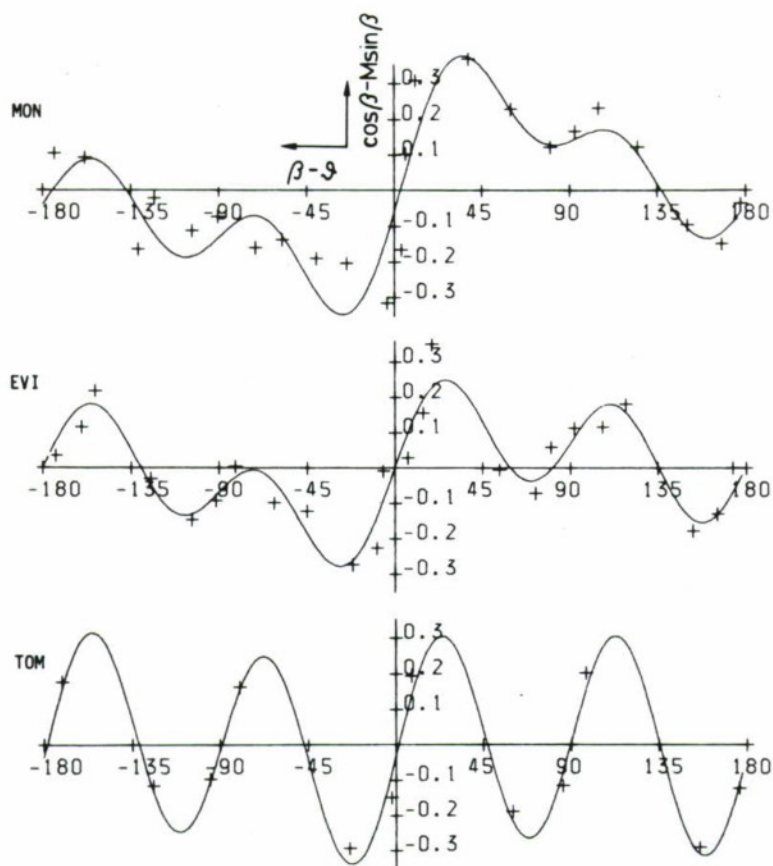
Figure 3.– Effect of the same tilted scene (a house) on the SV of three Ss (MON, EVI, TOM). The gravito-idiotropic torque –g is plotted as a function of $\beta - \vartheta$ (see Fig. 2). Crosses: means of pairs of settings. Curves: least-square fits of summed sine functions $-g = \sum V_n \sin n(\beta - \vartheta)$ with amplitudes $V_{1-4}$ to the data. Note the large variation of the amplitude $V_1$ of the first harmonic in contrast to the moderate variation of $V_2$ and $V_4$.

```
sin:                     p%
4 :0.203 +- 0.012       0.00
8 :0.055 +- 0.013       0.02
12:0.031 +- 0.010       0.18
cos:                     p%
4 :-0.032+- 0.008       0.06
8 :0.026 +- 0.010       2.06
12:0.008 +- 0.012      33.25
M:0.35      error:0.048
```
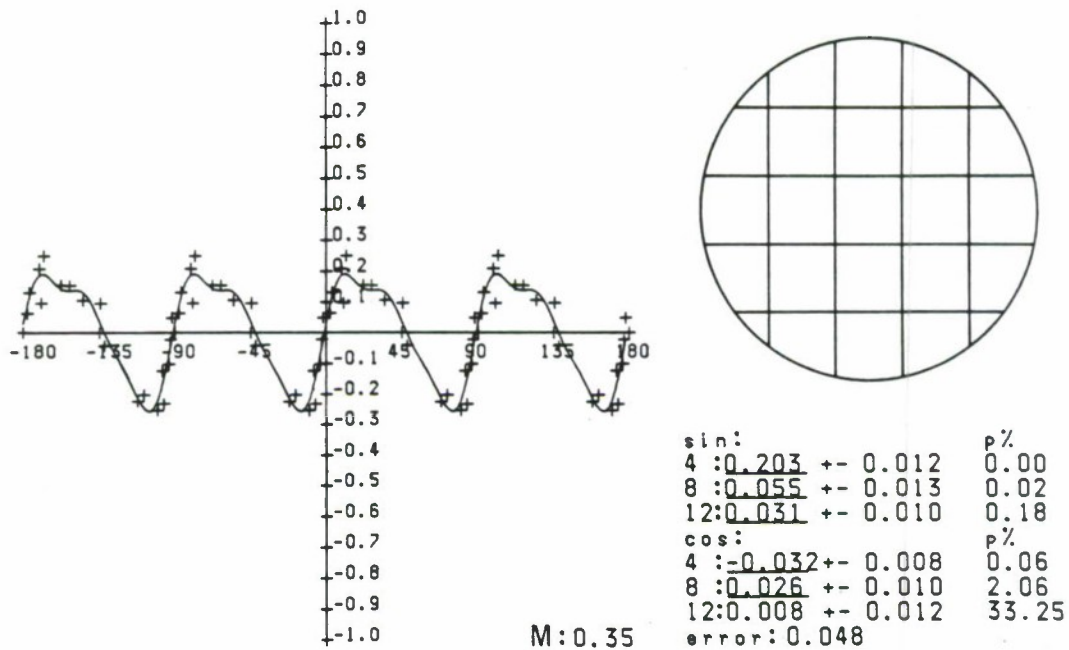
Figure 4.– Effect of pattern of squared luminous lines on SV. Method and evaluation as in Figs. 1-3. Inset gives numerical values of amplitudes (sines and cosines) of fourth, eighth, and twelfth harmonics of SV-function, their SD, and p (in %; two-tailed). Error means square deviation of data from approximation.

Romi Wilhelm 9.1.86 kl.Strich auf 'Sternenmann mit Armen'



```
sin:                     p%
1 :0.062 +- 0.015       0.06
2 :0.043 +- 0.015       0.70
3 :0.017 +- 0.015      15.45
4 :0.019 +- 0.015      12.41
cos:                     p%
1 :0.012 +- 0.015      26.97
2 :-0.020+- 0.015       9.86
3 :0.004 +- 0.014      40.79
4 :-0.002+- 0.015      44.44
M:0.49      error:0.042
```
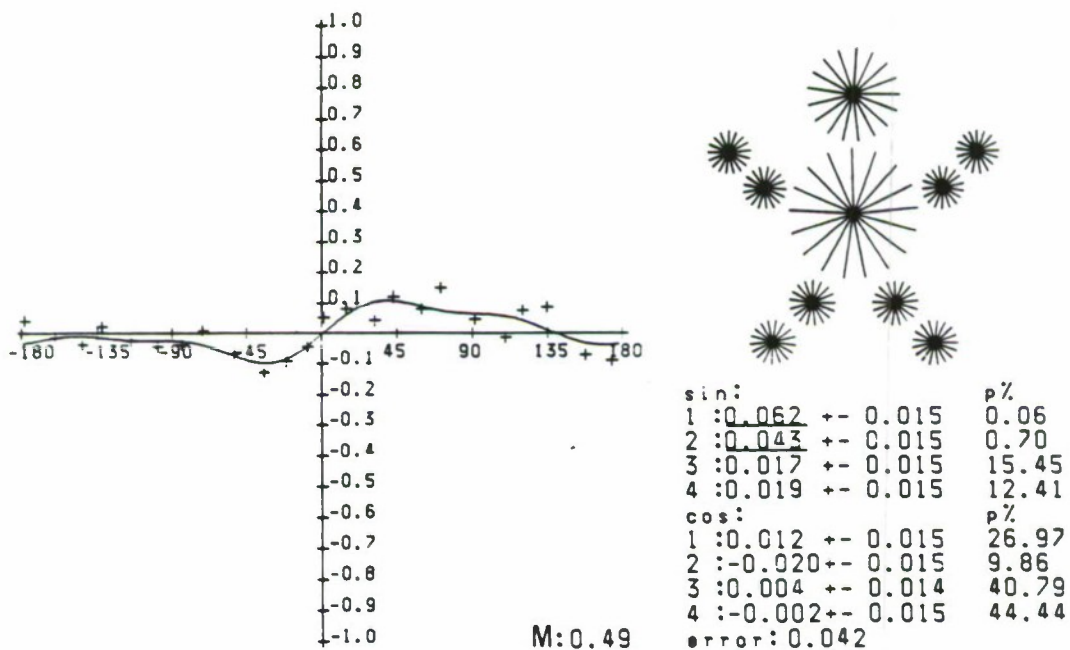
Figure 5.– Effect on SV of a figure which is composed of uniformly oriented luminous lines (star man). Procedures as in Figs. 1-3; symbols as in Fig. 4. Note that only the first and the second harmonic are significantly different from zero (two-tailed).

R.Wilhelm 3.12.85 kl.Strich auf 'Rautenmann'



```
sin:                  p%
1 :0.062 +- 0.024    1.26
2 :-0.011+- 0.020    30.32
3 :0.038 +- 0.024    7.01
4 :-0.147+- 0.024    0.00
cos:                  p%
1 :-0.023+- 0.024    31.76
2 :0.014 +- 0.027    31.72
3 :0.030 +- 0.024    13.41
4 :-0.002+- 0.023    48.02
error: 0.064
```

M:0.50
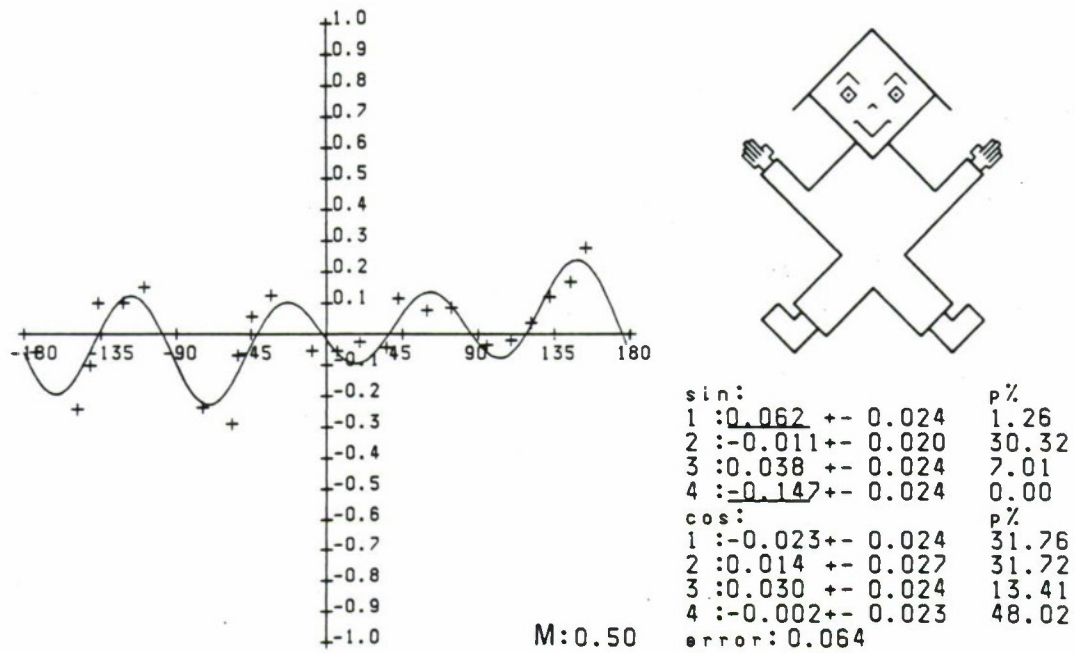
Figure 6.– Effect on SV of a figure which is composed of oblique luminous lines (diamond man). Note that the first and fourth harmonics are significantly (two-tailed) different from zero, but of different sign; that is, exactly (no cosines!) in counterphase at $\beta - \vartheta = 0$.
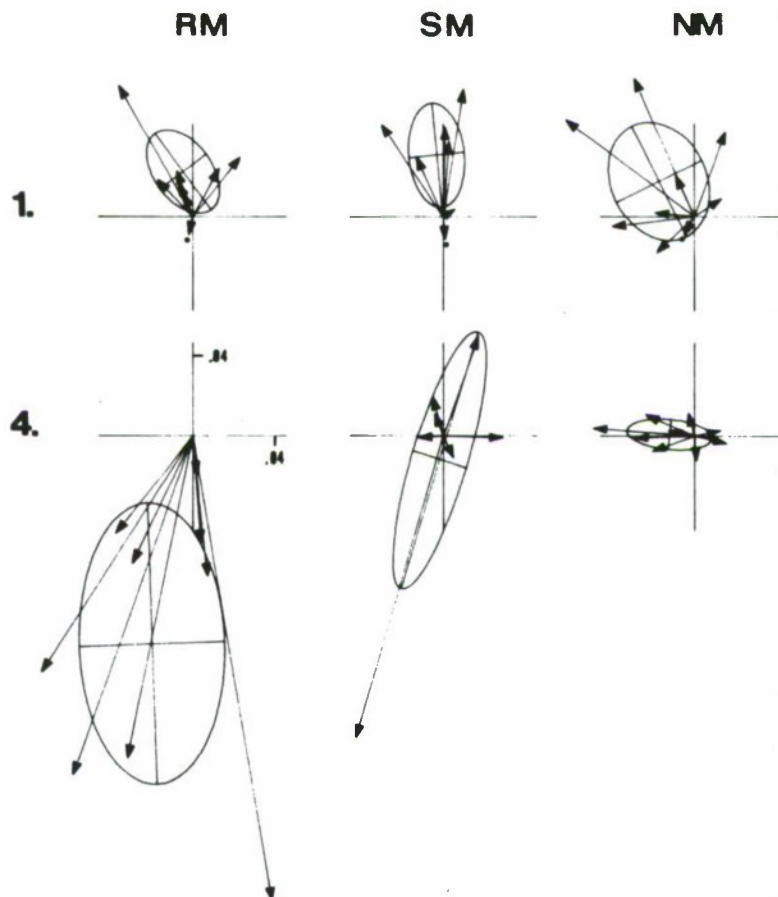
Figure 7.– First and fourth harmonics of experiments of Figs. 5 and 6 and in nine Ss. Location of arrowhead results from plotting sine amplitudes on ordinate and cosine amplitudes on abscissa (for scale see 4% marks on fourth of RM). RM: diamond man of Fig. 6; SM: star man of Fig. 5; NM figure in the shape of SM, but composed of randomly oriented dashes ("needle man"). Ellipses: two-dimensional SD. Note similarity of 1. harmonics for all figures and in all Ss except one (dot under arrowhead), who evinces a negative 1. harmonic, that is, sees the polarity inverted. Furthermore, only RM engenders a significant fourth harmonic.

42-13

Figure 8.– The well-known ambiguous figure appearing as witch or princess upon inversion of long axis.



|  | SALIENT | PREPONDERANT | AMBIGUOUS |
|---|---|---|---|
| PRINCESS | ● | ○ | ◉ |
| WITCH | ▲ | △ | ◉ |
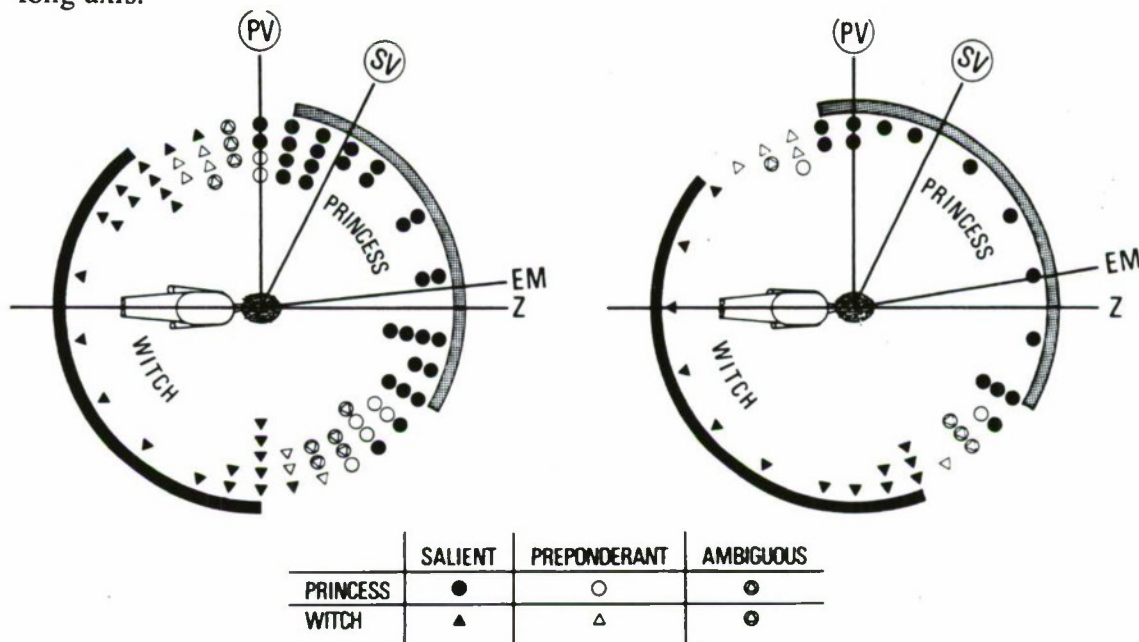
Figure 9.– To S lying on the side, the princess is presented in various static orientations. Direction of long (upright) axis *of princess* with respect to S's long axis (Z) is shown as direction of dot or triangle (like angle $\vartheta$ in Fig. 2). Type of symbol represents judgement of S on how the figure appears to her when presented in that direction. One symbol stands for one presentation. More presentations were made in directions of critical transitions than in those of complete salience (exclusive distinctness). The latter are connected by black (witch exclusive) or grey (princess exclusive) circular segments. Note that the direction of the midline of saliency coincides neither with direction of the SV nor with that of the vertical retinal meridian (RM), nor with that of the physical vertical (PV).

42-14

# OPTICAL, GRAVITATIONAL, AND KINESTHETIC DETERMINANTS OF JUDGED EYE LEVEL

Arnold E. Stoper and Malcolm M. Cohen
NASA Ames Research Center
Moffett Field, California

## SUMMARY

Subjects judged eye level, defined in three distinct ways relative to three distinct reference planes: 1) a gravitational horizontal, giving the "gravitationally referenced eye level" (GREL); 2) a visible surface, giving the "surface-referenced eye level" (SREL); and 3) a plane fixed with respect to the head, giving the "head-referenced eye level" (HREL). The information available for these judgments was varied by having the subjects view an illuminated target that could be placed in a box which: 1) was pitched at various angles, 2) was illuminated or kept in darkness, 3) was moved to different positions along the subject's head-to-foot body axis, and 4) was viewed with the subjects upright or reclining. Our results showed: 1) judgments of GREL made in the dark were 2.5° lower than in the light, with a significantly greater variability; 2) judged GREL was shifted approximately half of the way toward SREL when these two eye levels did not coincide; 3) judged SREL was shifted about 12% of the way toward HREL when these two eye levels did not coincide; 4) judged HREL was shifted about half way toward SREL when these two eye levels did not coincide and when the subject was upright (when the subject was reclining, HREL was shifted approximately 90% toward SREL); 5) the variability of the judged HREL in the dark was nearly twice as great with the subject reclining than with the subject upright. These results indicate that gravity is an important source of information for judgment of eye level. In the absence of information concerning the direction of gravity, the ability to judge HREL is extremely poor. A visible environment does not seem to afford precise information as to judgments of direction, but it probably does afford significant information as to the stability of these judgments.

## INTRODUCTION

A normal video display conveys fairly accurate information about exocentric directions among displayed visual objects (see Ellis, this volume), but not about egocentric directions, particularly those relative to eye level. This information is important to the observer in the natural environment, and can be used to advantage, especially in the case of a head-mounted display. The concern of the present paper is the mechanism underlying judgments of eye level, and the interactions of vision, gravitation, and bodily senses in these judgments.

There are at least three distinct meanings for visual eye level, all of which are important for the present analysis. Each meaning has associated with it a distinct reference plane with respect to which eye level can be specified. If a given reference plane passes through both the eye and a visual target, the target is said to be at that particular eye level. The three types of eye level are shown in figure 1, and described in table 1.

The Target/Head (T/H) system is responsible for the determination of the direction of a target relative to the head, or head-referenced eye level (HREL). This system presumably uses extra-retinal (e.g., kinesthetic or proprioceptive) eye position information (Matin, 1976). The Target/Gravity (T/G) system is responsible for the determination of the direction of a target relative to gravity, the gravitationally referenced eye level (GREL). It is composed of T/H and a Head/Gravity (H/G) system. The latter system presumably operates on the basis of vestibular (primarily otolithic) and postural information (Graybiel, 1973). The Target/Surface (T/S) system is responsible for determining the direction of a target relative to a visible surface, the surface-referenced eye level (SREL). In order to judge the direction of a target relative to the SREL, an observer must use optical information about the orientation of the surface; no extra-retinal, vestibular, or other proprioceptive information is necessary. The optical information involved might be in the form of depth cues which allow the observer to compare eye-to-surface distance with target-to-surface distance, or it might be in a form which allows a "direct" determination of SREL from optical information without recourse to judgments of distance (Gibson, 1950; Purdy, 1958; Sedgwick, 1980). Thus, in principle, T/S can be completely independent of T/H and T/G.

If an observer is standing on a level ground plane in a normal, illuminated, terrestrial environment, with head erect, all three eye levels (HREL, GREL, and SREL) coincide, and determination of any one automatically leads to determination of the other two. It is thus impossible, in that environment, to determine the relative contributions of the three physiological systems described. To do that, some means of separating them is necessary. Various methods to accomplish this separation were used in the following experiments.

## EXPERIMENT I: THE EFFECT OF ILLUMINATION ON JUDGMENT OF GREL

### Introduction and Method

Our experimental paradigm consisted simply in having the subject adjust a point of light to eye level, defined in one of the three ways above. First, we ask, "What contribution does optical information make to judgments of GREL?" To answer this question we simply turned off the lights. This eliminated optical information regarding orientation to the ground plane and all other environmental surfaces, and presumably eliminated information to the T/S system. The subject was seated in a dental chair which he or she could raise and lower hydraulically. (This technique minimized the possibility of the subject simply setting the target to the same visible point in each trial.) The task was to adjust the height of the chair so that the subject's eyes were "level" with a small target. (All three types of eye level are coincident in this situation.) A total of 80 trials occurred for each of 10 subjects.

### Results

Constant errors (which indicate accuracy) and standard deviations (which indicate precision) were calculated individually for each subject. The averages over all subjects are shown in table 2. The differences between light and dark are significant ($p < 0.01$ by ANOVA).

# DISCUSSION

The finding of higher constant error in the dark means that a small target appears to be about 2.5° higher in the dark than in the light. Others (MacDougall, 1903; Sandstrom, 1951) have found similar results. We have no satisfactory explanation for this effect.

The finding that eye level judgments are more variable in the dark is not surprising, nor is it easily explained. Three distinct hypotheses seem possible; the first two assume that T/S provides more accurate and precise directional information than T/G; the third makes no such assumption. The three hypotheses are

1. The "suppression" hypothesis assumes T/G is simply suppressed when T/S is available. If T/S is more precise than T/G, this suppression will result in improved precision.

2. The "weighted average" hypothesis assumes that the variability of the final judgment is a weighted average of the variabilities of T/G and T/S.

3. The "stability" hypothesis assumes that the function of optical information is to minimize the drift of directional judgments made by means of nonoptical information. Thus, no directional information per se is necessary from T/S, and no assumptions are made about its precision.

The following experiments are intended to help decide among these three hypotheses.

## EXPERIMENT 2: THE EFFECT OF PITCHED SURROUNDINGS ON GREL

### Introduction

Another way to study the interaction of the eye-level systems is to put them into "conflict." This effect has been extensively investigated in the roll dimension with the now classical "rod-and-frame" paradigm (Witkin and Asch, 1948).

### Method

A modification of the "pitchbox" method (Kleinhans, 1970) was used. Each of 12 subjects looked into a Styrofoam box, 30 cm wide by 45 cm high by 60 cm deep. The box was open at one end, and could be pitched 10° up or down (fig. 2).

Illumination was very dim (0.5 cd/m$^2$) to minimize visibility of surface features, but the inside edges of the box could be seen clearly. The apparatus allowed the pitchbox to be displaced linearly up or down as well as to be changed in pitch orientation. The subject could indicate eye level by adjusting the vertical position of a small target (produced by a laser beam).

In this experiment, the subject was instructed to set the target to the point in the pitchbox that was at his or her GREL. A 2x2x3x2 design with replication was used. The experiment consisted of four within-subject factors: (1) viewing condition (dark vs. light), (2) pitchbox position (high

vs. low: 6 cm apart), (3) pitchbox angle (10° up, level, or 10° down), and (4) laser starting position (up vs. down). Each factor combination was presented twice, yielding a total of 48 trials per subject.

## RESULTS AND DISCUSSION

### Box Pitch

Mean error of judged GREL is plotted in figure 3 as a function of orientation, position, and illumination of the pitchbox. It is clear that a strong effect of orientation on GREL exists in the light condition, but not in the dark. This can be described as a shift of judged GREL in the direction of true SREL. The magnitude of this shift is indicated by the slope of the judgment function. A total change in pitch (i.e., of SREL) of 20° produced a shift in GREL of 11.1° in the light, but only 1.5° in the dark. We will consider the slope of 0.55° (in the light) to be a measure of the strength of the effect of the visual environment. This effect is comparable in magnitude to that found by Matin and Fox (1986), and by Matin, Fox, and Doktorsky (1987). The simple fact of compromise between SREL and GREL means that T/G is not totally suppressed, even while T/S is operating, and is strong evidence against the suppression hypothesis.

### Box Height

The effect of box height is clearly evident in the figure. The linear shift of the pitchbox of 6 cm (5.5° of visual angle) produced a 1.47 cm (1.35°) shift in GREL. This is comparable in magnitude to a similar linear displacement effect found by Kleinhans (1970). It may be due to the Dietzel-Roelofs effect (Howard, 1982, p. 302), where the apparent straight ahead is displaced toward the center of an asymmetrical visual display. Another possible explanation is a tendency for subjects to set eye level toward the same optically determined point on each successive trial. Whatever the cause of this effect, it may account for as much as 40% of the orientation effect, since with our apparatus, a change in orientation also produced a displacement of the visual scene.

### Variability

It might be expected that conflict between two systems would greatly increase variability. For example, each system could contribute a component equal to its own variability, and there would be an additional component caused by variability in combining the systems. Figure 4 shows within-subject standard deviations calculated separately for each of the three orientations, in the light and the dark.

Here it can be seen that variability of judgment in the dark is higher than in the light; however, it is not affected by orientation. There is no more variability when the systems are in conflict (at ±10°) than when they are not (when the pitchbox is level, at 0°). This finding indicates that the weighting of the systems is very stable over a series of trials for each subject.

# EXPERIMENT 3. THE EFFECT OF GRAVITY ON SREL JUDGMENTS

## Introduction and Method

To observe the operation of T/S, we instructed the subject to align his or her line of sight with the floor of the movable pitchbox, thus judging the SREL. Just as we "turned off" T/S by extinguishing the light, we can turn off T/G by orienting the subject so that gravity does not abet the task. Each of 12 subjects judged SREL, both with upright posture, when they could presumably use gravitational information and T/G, and reclining on the left side, where gravity and T/G were of no use. (The T/H system presumably continued to operate in both conditions.) In the upright condition the method was identical to that of Experiment 2, except that the instructions were to find SREL rather than GREL. In the reclining condition the entire apparatus (shown in fig. 2) was rotated 90°.

As in Experiment 2, the pitchbox was set in two different positions displaced 6 cm along the subject's longitudinal body axis (Z axis).

## Results and Discussion

Results are plotted in figures 5 and 6. ANOVA showed significant effects of box pitch and box height.

## Box Pitch

There is a clear shift of SREL judgments in the direction of HREL in both the upright and reclining conditions. The slope is 0.15, much less than the 0.55 found in Experiment 2. (Note that, while Experiment 2 showed an effect of optical variables on a nonoptical judgment, the present experiment found an effect of nonoptical variables on an optical judgment.) The fact that the slope is essentially the same for both upright and reclining body orientations implies that T/H rather than T/G is producing the bias we obtained. This result is similar to that of Mittlestaedt (1983).

## Box Height

The effect of the 6-cm box displacement was a shift of 2.47 cm (2.26°) in the upright and 3.5 cm (3.21°) in the reclining condition. The size of this effect implies that the subjects did not effectively use the optical orientation information available to them. Instead, they seem to have had a strong tendency to set the target near the same location on the back of the box with each trial.

## Variability

Standard deviations for SREL judgments are shown in figure 7.

SREL judgments made with the subject upright showed greater within-subject variability than those made with the subject reclining. This observation may be taken to imply that gravity does not enhance the precision of SREL judgments under upright conditions.

## EXPERIMENT 4: THE EFFECT OF GRAVITY AND PITCHED SURROUNDINGS ON HREL JUDGMENTS

### Introduction and Method

To observe the influence of T/S on T/H, we instructed the subject to set his or her eyes "straight ahead" and place the target at the fixation point, thus judging HREL. In the upright condition the method was identical to that of Experiment 2, except that the instructions were to find HREL rather than GREL. The reclining condition arrangement was identical to that of Experiment 3.

### Results and Discussion

Results are plotted in figures 8 and 9. ANOVA showed significant effects of orientation and box height.

### Box Pitch

There is a clear shift of HREL judgments in the direction of SREL in both the upright and reclining conditions. The slope for the judgments of HREL with upright posture in the light is 0.45, about the same magnitude as was observed in Experiment 2. We thought that this effect could be due to a confusion of instructions when HREL and GREL were coincident, and we expected a much weaker effect in the reclining conditions, when GREL was absent. In fact, however, a much stronger effect was found (slope = 0.89). This can be explained in terms of Mittlestaedt's (1986) vector combination model. In the upright condition, both T/G and T/H indicate a more or less horizontal eye level, and T/S would be combined with both of these. In the reclining condition T/S combines with only T/H. The result in the reclining condition is thus closer to T/S.

### Variability

It can be seen in figure 10 that, for upright posture, the variabilities of HREL and GREL judgments are very similar, both in the dark and in the light. For reclining posture, however, HREL variability is twice as great in the dark as in the light. This result indicates that the presence of gravitational information has a stabilizing effect on HREL judgments.

# CONCLUSIONS

1. Increased precision in the light. We present evidence against both the suppression and the weighted average hypotheses. Only the stability hypothesis is not contradicted by these data. This hypothesis could be tested directly by using a random dot field as a visual environment. Such a field would have no direction information, so any improvement in precision of GREL would be by. means of stability information.

2. Box displacement effect. This may be a significant factor in the orientation effect. It could be controlled in a future experiment by rotating the pitchbox around the center of its back, rather than around the subject's eye.

The large size of this effect when judging SREL indicates that ability to judge orientation of the line of sight in the pitch dimension relative to a surface on the basis of purely optical information is poor under the conditions of this experiment.

3. Head relative information. Perhaps our most surprising result was the almost complete "visual capture" of HREL judgments in the light while the subject was reclining on his or her side in Experiment 4, and the corresponding high variability of these judgments in the dark. Both of these results indicate very low ability to use T/H to judge eye level in the absence of gravity information. In more practical terms, this result indicates that judgment of the pitch of the observer's head (and by implication, the rest of his or her body) relative to a surface is much less precise, and subject to a much higher degree of visual capture, when gravity is not present to aid this judgment.

# REFERENCES

Ellis, S. (1987). Pictorial Communication: Pictures and the Synthetic Universe. Presentation to Spatial Displays and Spatial Instruments Conference: Asilomar, California.

Gibson, J. J. (1950). The perception of the visual world. Boston: Houghton Mifflin.

Graybiel, A. (1973). The Vestibular System. In: Bioastronautics Data Book. NASA SP 3006, 533-609.

Howard, I. P. (1982). Human visual orientation. New York: Wiley.

Kleinhans, J. L. (1970). Perception of spatial orientation in sloped, slanted and tilted visual fields. Ph.D. Dissertation, Rutgers Univ., New Jersey.

MacDougall, R. (1903). The subjective horizon. Psychol. Rev. Monogr. Supp., 4, 145-166.

Matin, L. (1976). Saccades and extraretinal signal for visual direction. In R. A. Monty and J. W. Senders (Eds.), Eye Movements and Psychological Processes. New Jersey: Erlbaum, 205-220.

Matin, L. and Fox, C. R. (1986). Perceived eye level: Elevation jointly determined by visual field, pitch, EEPI, and gravity. Investigative Ophthalmol. Visual Sci. (Supp.) 27, 333.

Matin, L., Fox, C. R., and Doktorsky, Y. (1987). How high is up? II. Investigative Ophthalmol. Visual Sci. (Supp.), 28, 300.

Mittlestaedt, H. (1986). A new solution to the problem of the subjective vertical. Naturwissenschaften, 70, 272-281.

Purdy, W. C. (1958). The hypothesis of psychophysical correspondence in space perception. Doctoral Dissertation, Cornell Univ. (Ann Arbor: University Microfilms, No. 58-5594).

Sandstrom, C. I. (1951). Orientation in the present space. Upsala: Almqvist and Wicksell.

Sedgwick, H. A. (1980). The geometry of spatial layout in pictorial representation. M. Hagen (Ed.), The perception of pictures (pp. 33-88). New York: Academic Press.

Stoper, A. E., and Cohen, M. M. (1986). Judgments of eye level in light and darkness. Percep. Psychophys., 40, 311-316.

Witkin, H. A. and Asch, S. E. (1948). Studies in space orientation. IV. Further experiments on perception of the upright with displaced visual fields. J. Exp. Psychol., 38, 762-782.

## TABLE 1. TYPES OF EYE LEVEL

| Symbol | Type | Physiological system | Reference plane |
|--------|------|----------------------|-----------------|
| HREL | Head-referenced eye level | Target/head (T/H) | Arbitrary plane tied to head |
| GREL | Gravity-referenced eye level | Target/gravity (T/G) (T/G = T/H + H/G) | Gravitational horizontal |
| SREL | Surface-referenced eye level | Target/Surface (T/S) | Ground surface or other visible plane surface |


## TABLE 2. MEANS AND STANDARD DEVIATIONS (DEG) FOR ERROR IN EYE-LEVEL JUDGMENTS IN LIGHT AND DARK, Average of 10 subjects (Stoper and Cohen, 1986)

|  | Light | Dark |
|--|-------|------|
| Constant error (mean) | 0.29 | 2.79 |
| Variable error (standard deviation) | 1.03 | 1.72 |
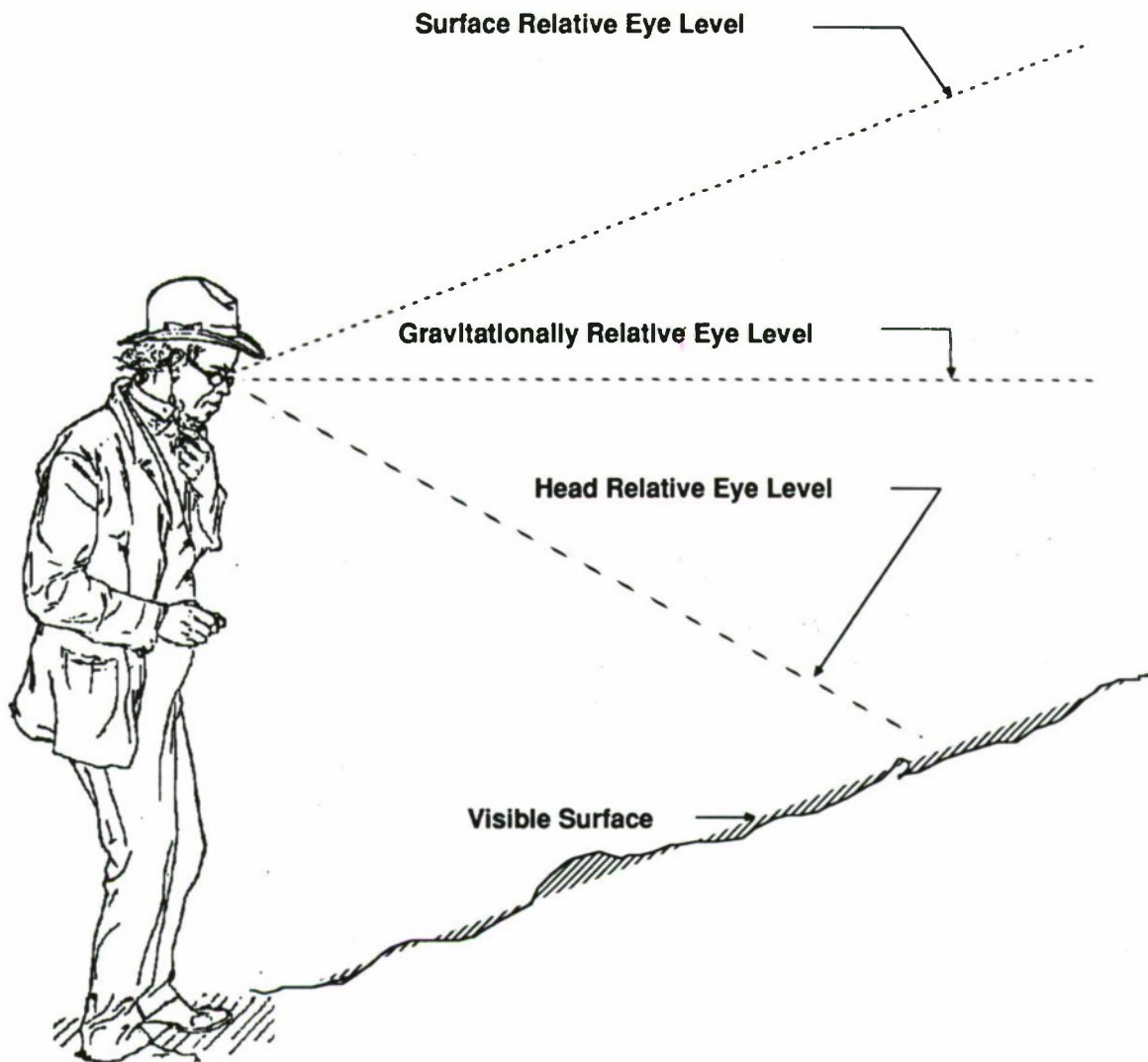
Figure 1.– Three types of eye level in normal terrestrial environment.  See table 1 for description.
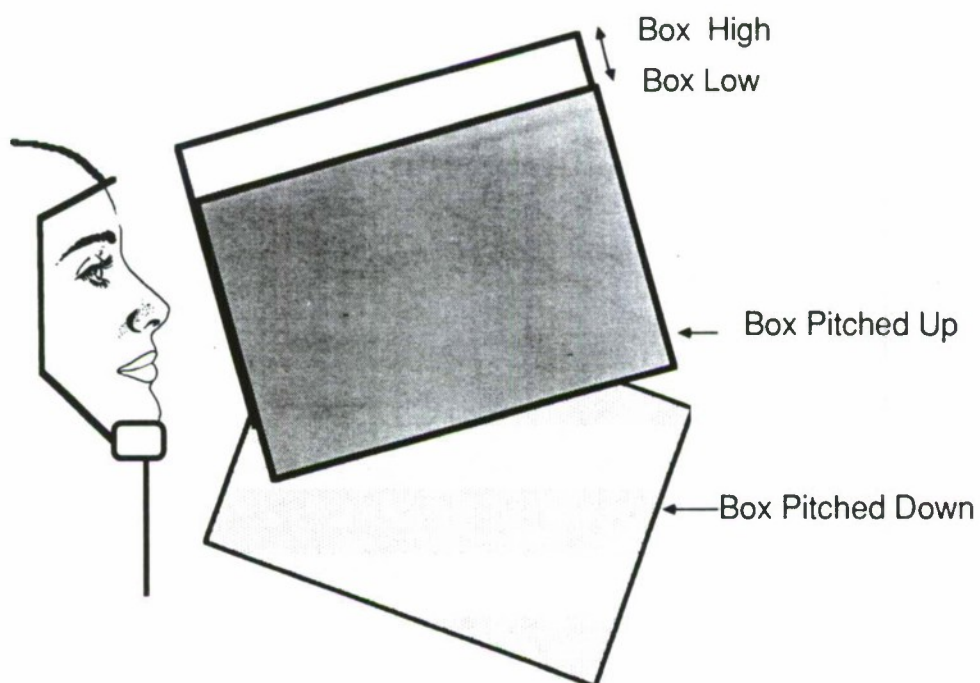
Figure 2.– Orientations and positions of the pitchbox.
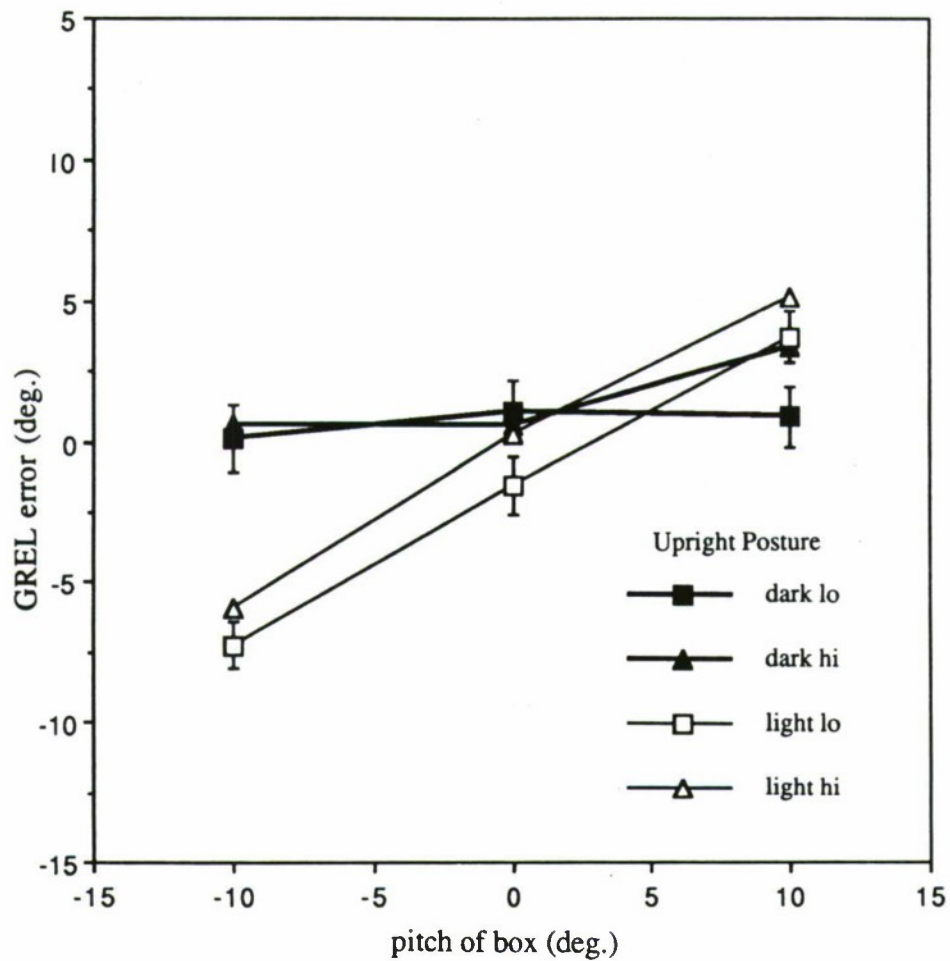
Figure 3.– Mean error in judgment of gravitationally relative eye level (GREL) of 12 subjects as a function of orientation, position, and illumination of the pitchbox. Pitch of +10° means the pitchbox was pitched up. Error bars represent the standard error of the mean (between subjects).
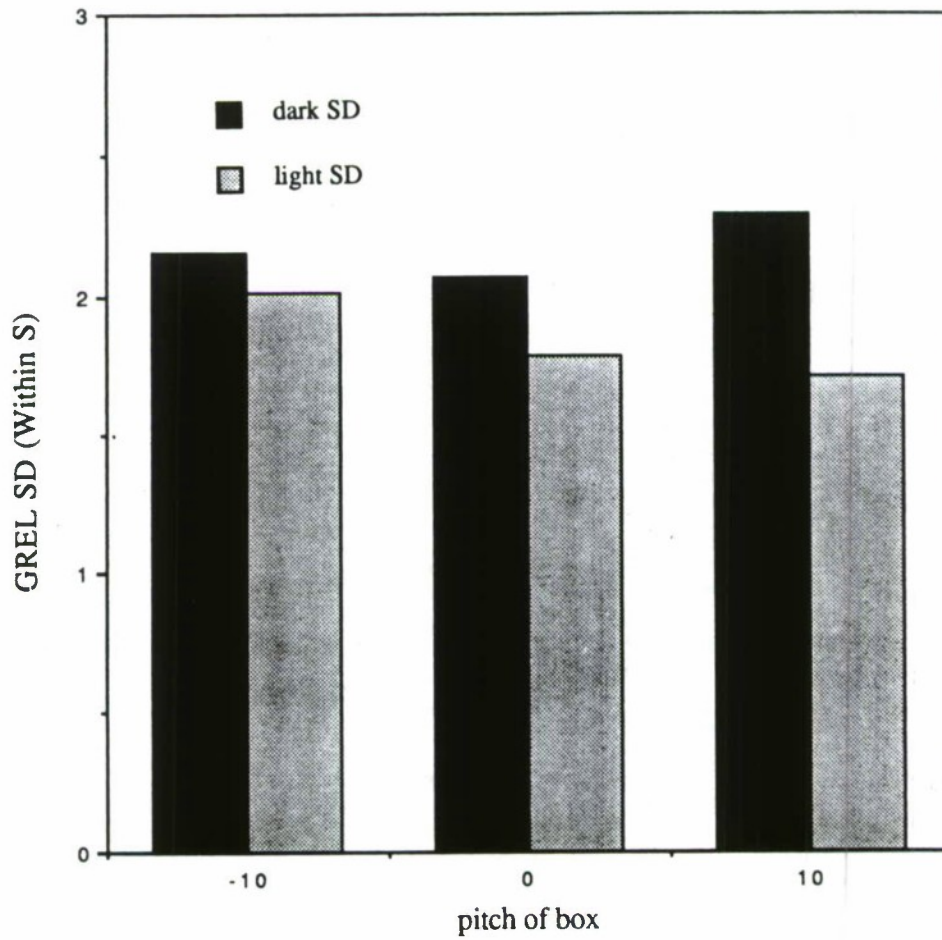
Figure 4.– Standard deviations (within subjects) of GREL judgments of 12 subjects for each of three orientations, in the light and in the dark.
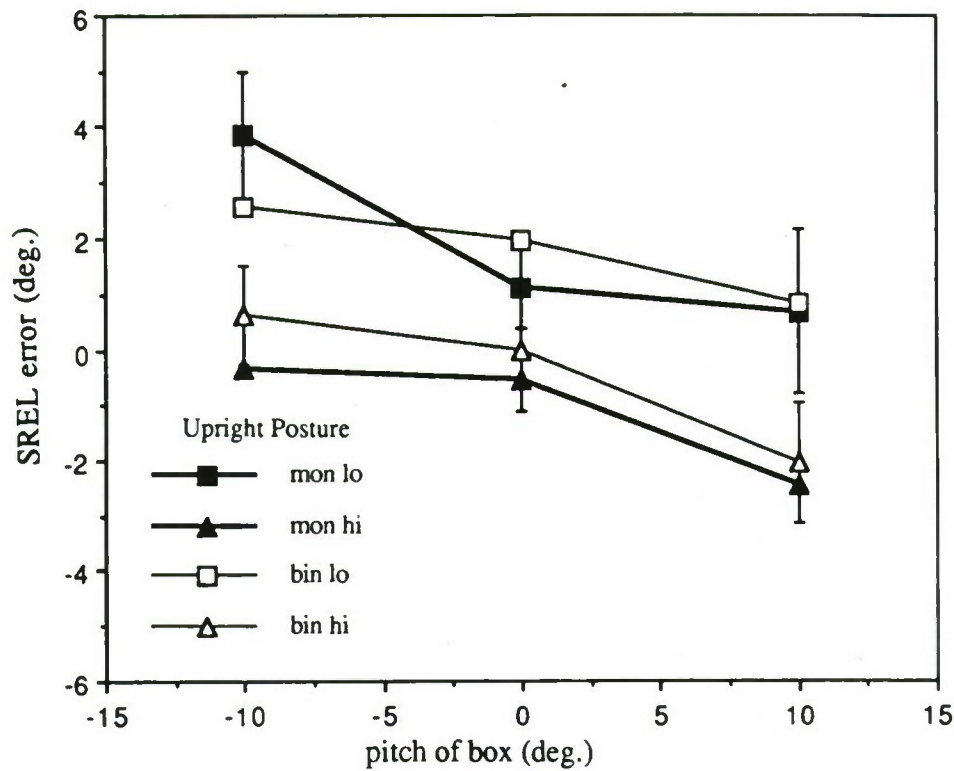
Figure 5.– Mean error in judgment of surface-relative eye level (SREL) of 12 subjects as a function of orientation, position, and illumination of the pitchbox; judgments made with upright posture. Error bars represent the standard error of the mean (between subjects).
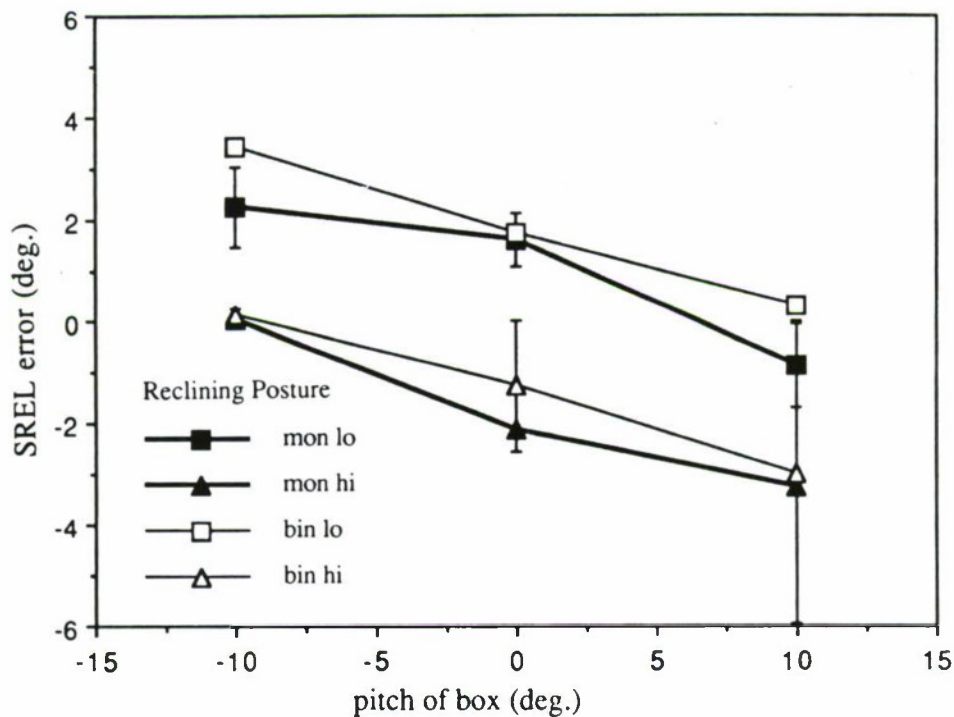


Figure 6.– Mean error in judgment of SREL of 12 subjects; judgments made with reclining posture.
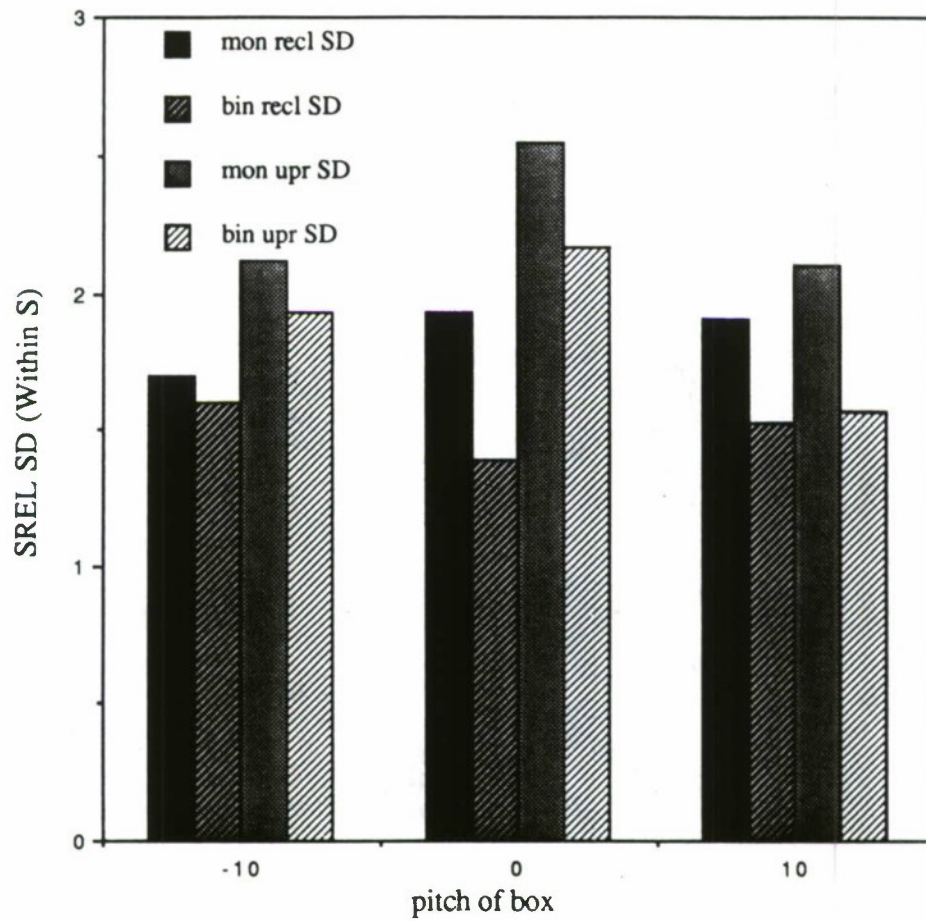
Figure 7.– Standard deviations (within subjects) of SREL judgments of 12 subjects for each of three orientations, in the light and in the dark.
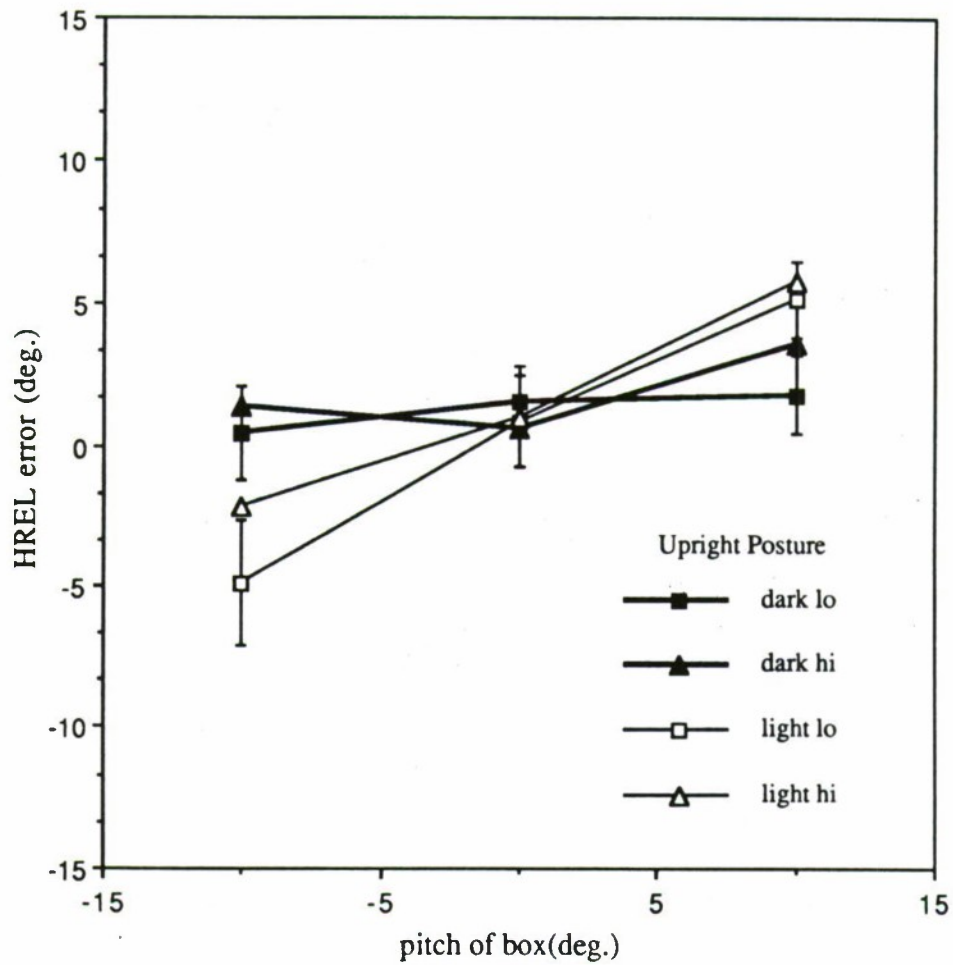
Figure 8.– Mean error in judgment of head-relative eye level (HREL) of 12 subjects as a function of orientation, position, and illumination of the pitchbox; judgments made with upright posture. Error bars represent the standard error of the mean (between subjects).
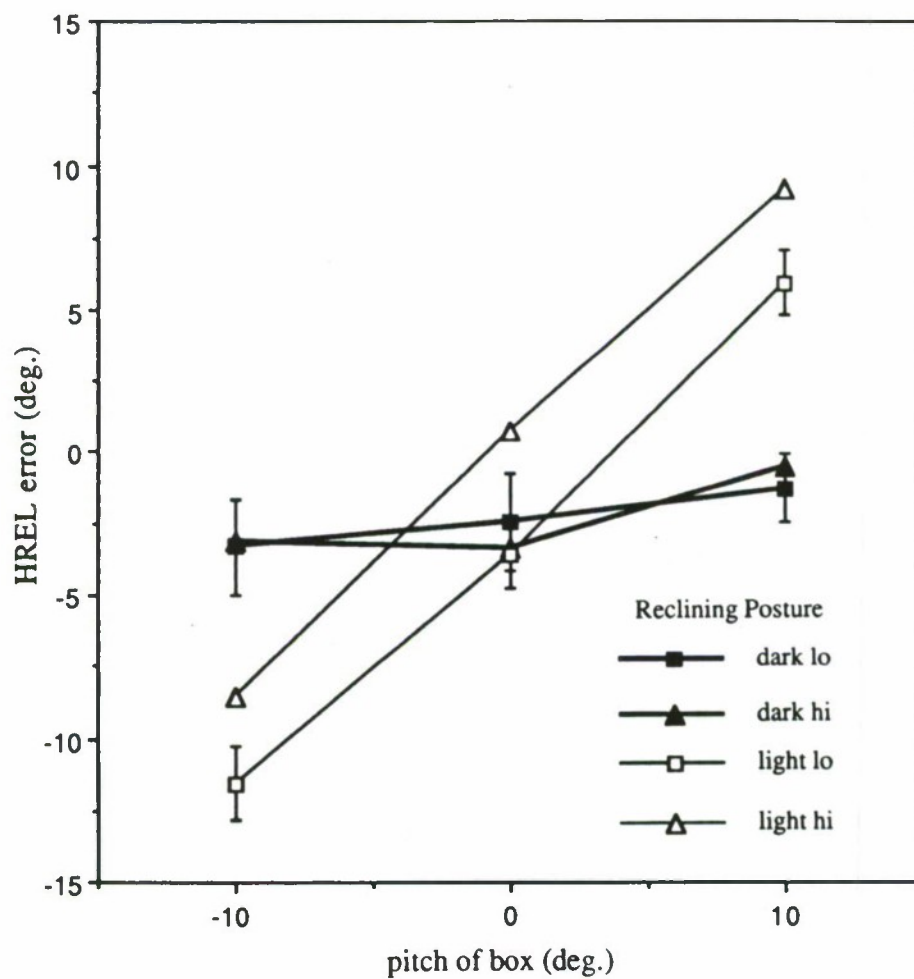
Figure 9.– Mean error in judgment of HREL of 12 subjects; judgments made with reclining posture.
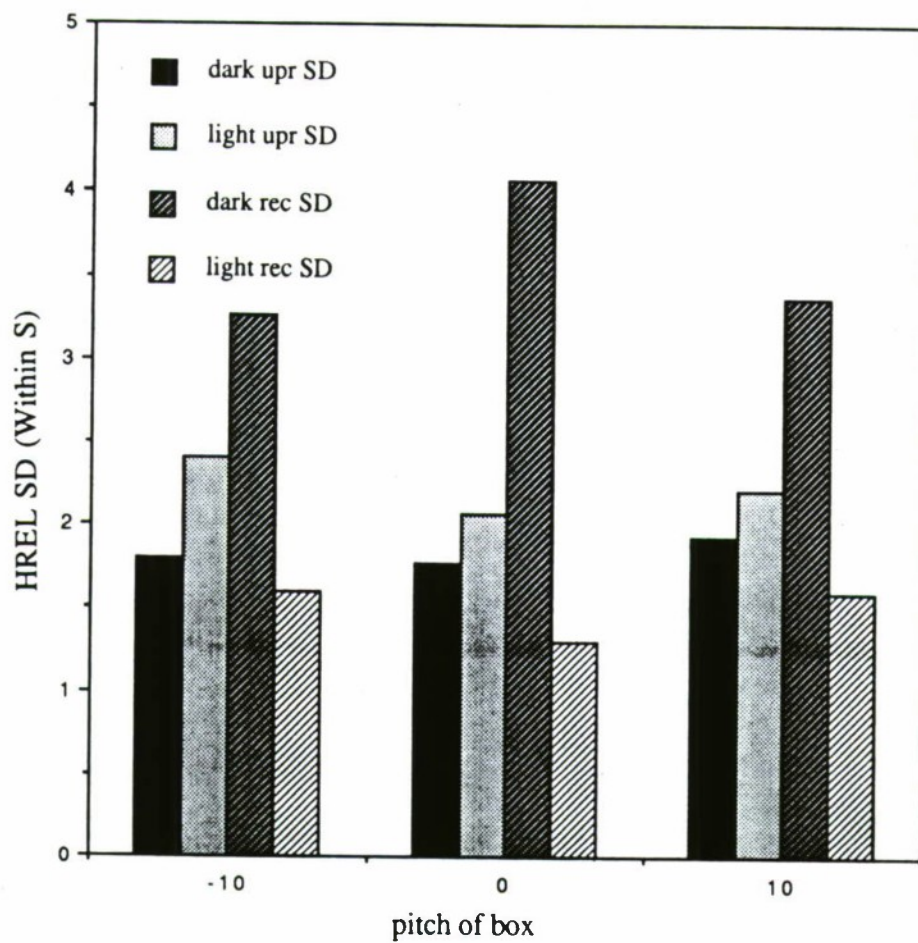
Figure 10.– Standard deviations (within subjects) of HREL judgments of 12 subjects for each of three orientations, in the light and in the dark.

# VOLUNTARY PRESETTING OF THE VESTIBULAR OCULAR REFLEX PERMITS GAZE STABILIZATION DESPITE PERTURBATION OF FAST HEAD MOVEMENTS

Wolfgang H. Zangemeister
Hamburg University Neurological Clinic
Hamburg, West Germany

## SUMMARY

Normal subjects are able to change voluntarily and continuously their head-eye latency together with their compensatory eye movement gain. A continuous spectrum of intent-latency modes of the subject's coordinated gaze through verbal feedback could be demonstrated. It was also demonstrated that the intent to counteract any perturbation of head-eye movement, i.e., the mental set, permitted the subjects to manipulate consciously their vestibular ocular reflex (VOR) gain. From our data we infer that the VOR is always "on." It may be, however, variably suppressed by higher cortical control. With appropriate training, head-mounted displays should permit an easy VOR presetting that leads to image stabilization, perhaps together with a decrease of possible misjudgments.

## INTRODUCTION

For some time it has been known that visual and mental effort influence the vestibular ocular reflex (VOR). Besides visual long- and short-term adaptation to reversing prisms (Melvill Jones and Gonshor, 1982) and fixation suppression of the VOR (Takemori and Cohen, 1974; Dichgans et al., 1978; Zangemeister and Hansen, 1986), the mental set of a subject can influence the VOR, e.g., through an imagined target (Barr et al., 1976; Melvill Jones et al., 1984) or anticipatory intent only (Zangemeister and Stark, 1981). In contrast to animals, human head and eye movements are governed by a conscious will of the human performer that includes verbal communication. Thus in a given experimental setup, the synkinesis of active human gaze may be changed according to instruction. The verbal feedback to the subject might permit a whole range of gaze types, even with amplitude and prediction of a visual target being constant. The gaze types (Zangemeister and Stark, 1982a) are defined by head minus eye latency differences (table 1). This has been demonstrated particularly by looking at the timing of the neck elektromyogram as the head movement control signal (Zangemeister et al., 1982b; Zangemeister and Stark, 1983; Stark et al., 1986). In this study, we compared the voluntarily changeable human gaze types performed during the same experiment with and without the addition of a randomly applied perturbation to the head-eye movement system. We tried to answer three questions in particular:

1. Are we able to modulate continuously the types of coordinated gaze through conscious intent during predictive active head movements?

2. What is the gaze (saccade and VOR/CEM (compensatory eye movement)) response to passive random head rotation from zero head velocity with respect to the preset intent of a given subject?

3. Does random perturbation of the head during the·early phase of gaze acceleration generate responses that are the sum of responses to experiment (1) and (2)?

## METHODS

Eye movements were recorded by monocular DC Electrooculography, head movements by using a horizontal angular accelerometer (Schaevitz) and a high-resolution ceramic potentiometer linked to the head through universal joints (Zangemeister and Stark, 1982c). Twelve normal subjects (age 22-25) attended a semicircular screen sitting in a darkened room. While they actively performed fast horizontal (saccadic) head rotations between two continuously lit targets at ±30° amplitude with a frequency around 0.3 Hz, they were instructed to focus on the following tasks: (1) "shift your eyes ahead of your head," (2) "shift your head ahead of your eyes." During (1) they were instructed to shift eyes "long before" (i, type II), or "shortly before" (ii, type I) the head. During (2) they were instructed to shift head "earlier" (i, type IIIA), or "much earlier" (II, type IIIB) than the eye, eventually "with the intent to suppress any eye movement" (type IIIB or IV). Each task included 50 to 100 head movements.

Perturbations were done pseudorandomly, (1) from a zero P,V,A (position, velocity, acceleration) initial condition of the head-eye movement system, and (2) during the early phase of head acceleration. They consisted of (1) fast passive head accelerations, of (2) short decelerating or accelerating impulses during the early phase of active head acceleration and were recorded by the head-mounted accelerometer. Perturbation impulses were generated through an apparatus that permitted manual acceleration or deceleration of the head through cords that were tangentially linked directly to the tightly set head helmet.

## RESULTS

1. The subjects demonstrated their ability (fig. 1) to switch between gaze types in the experimentally set predictive situation of constant and large-amplitude targets. The respective gains (eye/head velocity) were: ty.II 0.9-1.1, ty.III 0.13, ty.IV 0.06-0.09. This result was expected from our earlier studies (Zangemeister and Huefner, 1984; Zangemeister and Stark, 1982a,c). The subjects showed differing amounts of success in performing the intended gaze type, with type IV being the most difficult to perform, supposedly because of the high concentration necessary (table 1).

2. Random perturbation of the head while in primary position, with head velocity and acceleration being zero (fig. 2), resulted in large saccades/quick phases of long duration, and a large and delayed VOR/CEM, if the subject had low preset intent to withstand the perturbation; in this case head acceleration showed a long-lasting damped oscillation. Respective gains were: figure 2 (upper): 0.35 (upper) 0.45 (lower); figure 2 (lower left): 0.5 (upper), 0.17 (lower). With increasing intent of the subject (fig. 2 left, middle, and lower), head acceleration finally became highly overdamped, but still with comparable initial acceleration values, and eye movements showed increasingly smaller and shorter quick phases as well as an early short VOR response. In addition, with the highest intent a late anticompensatory eye movement was obtained.

3a. Random perturbations of the accelerating head, i.e., sudden acceleration or deceleration of gaze in flight (fig. 3), were characterized by small VOR responses after the perturbation in case of high intent of the subject as in gaze type IIIB, or much higher VOR/CEM gain in case of low intent comparable to gaze type I. Respective gains were: figure 3 (left) ty.I 0.55, ty.3 0.06, ty.IV 0.08 (left), 0.09 (right); figure 3 (right): 0.13 (upper), 0.90 (lower).

3b. Random perturbations were also applied during coordinated head-eye movements in pursuit of a sinusoidally moving target (maximum velocity 50°/sec) with the VOR being suppressed through constant fixation of the pursuit target. Figure 4 (left) demonstrates the different amount of VOR fixation suppression as a function of changing intent during fixation of a sinusoidal target of the same frequency. With perturbation (fig. 4, right) a response was obtained that was comparable to the result of experiment (2). That is, depending on the subject's intent and concentration, the VOR response was low for high intent and vice versa (gain fig. 4, right: 0.044).

Therefore, the three initial questions could be answered as follows:

1. In nonrandom situations subjects can intentionally and continuously change their gaze types.

2. Gaze responses to passive random head accelerations depend on the subject's preset intent.

3. Perturbation of predictive gaze saccades in midflight results in the sum of tasks one and two.

## DISCUSSION

The input-output characteristics of the VOR are subject to major moment-to-moment fluctuations depending on nonvisual factors, such as state of "arousal" (Melvill Jones and Sugie, 1972) and mental set (Collins, 1962). More recently, it has been found that the influence of "mental set" depends explicitly upon the subject's conscious choice of intended visual goal (Barr et al., 1976; Sharpe et al., 19081; Baloh et al., 1984; Fuller et al., 1983), i.e., following earth-fixed or head-fixed targets during head rotation. Consistent alteration of the mentally chosen goal can alone produce adaptive alteration of internal parameters controlling VOR gain (Berthoz and Melvill Jones, 1985). Obviously, comparison of afferent retinal slip detectors with concurrent vestibular afferents can be substituted by a "working" comparison made between the vestibular input and an efferent feedback copy of either the concurrent, or the imagined or anticipated concurrent, oculomotor output, as proposed by Miles and Eighmy (1980).

Our results here demonstrate the ability of the subjects to perform short-term adaptation during verbal feedback instructing for eye-head latency changes that changed the types of active gaze. These results are comparable to the data from Barr et al. (1976), in that an almost immediate change between different VOR gains with constant visual input could be generated. In addition, our perturbation experiments expanded these data, demonstrating the task- (or gaze-type) dependent attenuation of the VOR. This is in contrast to results in animals, where perturbation of

visually triggered eye-head saccades resulted in an acceleration of the eye (Guitton et al., 1984; Fuller et al., 1983), because a conscious task-influence of the VOR is impossible. Therefore not only can a representation of the target's percept (Barr et al., 1976) be created, but also an internal image of the anticipated VOR response in conjunction with the appropriate saccade.

We hypothesize that through the cortico-cerebellar loop a given subject is able to continuously eliminate the VOR response during predictive gaze movements. This is done internally by generating an image of the anticipated VOR response in conjunction with the appropriate saccade, and then subtracting it from the actual reflex response. This internal image can be manipulated intentionally and continuously WITHOUT a VOR on/off switch. In this way a flexible adaptation of the conscious subject to anticipated tasks is performed.

# REFERENCES

Baloh, R. W., Lyerly, R., Yee, R. D., and Houmstra, V. (1984) Voluntary control of the human vestibulo-ocular reflex. Acta Otolaryngol. 97:1-6.

Barr, C. C., Schultheis, L., and Robinson, D. A. (1976) Voluntary, non-visual control of the human vestibulo-ocular reflex. Acta Otolaryngol. 81:365-375.

Berthoz, A., Melvill Jones, G. (1985) Adaptive Mechanisms in Gaze Control. Amsterdam-New York.

Collins, W. E. (1962) Effects of mental set upon vestibular nystagmus. J. Exp. Psychol. 63:191-197.

Dichgans, J., Reutern, V. Reutern, G., and Römmelt, U. (1978) Impaired suppression of vestibular nystagmus by fixation. Arch. Psychiatr. Nervenkr. 226:183-199.

Fuller, J. H., Maldonado, H., and Schlag, J. (1983) Vestibular-oculomotor interaction in cat eye-head movements. Exp. Brain Res. 271:241-250.

Guitton, D., Douglas, R., and Volle, M. (1984) Eye-head coordination in cats. J. Neurophysiol. 52:1030-1050.

Melvill Jones, G., Sugie, N. (1972) Vestibulo-ocular responses during sleep in man. EEG Clin. Neurophysiol. 32:43-53.

Melvill Jones, G., and Gonshor, A. (1982) Oculomotor response to rapid head oscillation after prolonged adaptation to vision reversal. Expt. Brain Res. 45:45-58.

Melvill Jones, G., Berthoz, A., and Segal, B. (1984) Adaptive modification of the vestibulo-ocular reflex by mental effort in darkness. Exp. Brain Res. 56:149-153.

Miles, F. A. and Eighmy, B. B. (1980) Long-term adaptive changes in primate vestibulo-ocular reflex. J. Neurophysiol. 43:1406-1425.

Sharpe, J. A., Goldberg, H. J., Lo, A. W., and Hensham, Y. O. (1981) Visual-vestibular interaction in multiple sklerosis. Neurology 31:427-433.

Stark, L., Zangemeister, W. H., Hannaford, B., and Kunze, K. (1986) Use of models of brainstem reflexes for clinical research. In: Clinical problems of brainstem disorders, ed. by K. Kunze, W. H. Zangemeister, A. Arlt, Stuttgart-New York, pp. 172-184.

Takemori, S., and Cohen, B. (1974) Loss of visual suppression of vestibular nystagmus after flocculus lesions. Brain Res. 72:213-224.

Zangemeister, W. H., and Stark, L. (1981) Active head rotation and eye-head coordination. In: Vestibular and Oculomotor Physiology, ed. by B. Cohen. Ann. N.Y. Acad. Sci. 374:540-559.

Zangemeister, W. H., and Stark, L. (1982a) Gaze types: interactions of eye and head movements in gaze. Exp. Neurol. 77:563-577.

Zangemeister, W. H., Stark, L., Meienberg, O., and Waite, T. (1982b) Motor control of head movements. J. Neurol. Sci. 55:1-14.

Zangemeister, W. H., and Stark, L. (1982c) Gaze latency: variable interactions of eye and head in gaze. Exp. Neurol. 75:389-406.

Zangemeister, W. H., and Stark, L. (1983) Pathological types of eye and head gaze coordination. Neuro-ophthalmol. 3:259-276.

Zangemeister, W. H., and Huefner, G. (1984) Saccades during active head movements: interactive gaze types. In: Theoretical and Applied Aspects of Eye Movement Research, ed. by A. Gale and F. Johnson, Amsterdam-New York, pp. 113-122.

Zangemeister, W. H., and Hansen, H. C. (1986) Fixation suppression of the vestibular ocular reflex and head movement correlated EEG potentials. In: Eye Movements, ed. by J. K. O'Reagan and A. Levy-Schoen, Amsterdam-New York, pp. 247-256.

Table 1.– Gaze types defined by latency: eye minus head latency. Type II: early prediction of eye, late head movement; eye movement dominates gaze. I: head follows eye shortly before eye has reached target; classical gaze type. III: head and eye movements start about simultaneously. Predictive gaze type. IV: early prediction of head, late eye saccade; head movement dominates gaze. Suppression of VOR/CEM in III and IV. See also figure 1b.

| Type | Eyelatency-headlatency, msec | Average rate of success in generating intentionally different gaze types through verbal feedback,% |
|------|------------------------------|--------------------------------------------------------------------------|
| I    | +50                          | 76                                                                       |
| II   | <50                          | 56                                                                       |
| IIIa | >50-200                      | 69                                                                       |
| IIIb | >200-550                     |                                                                          |
| IV   | >550                         | 16                                                                       |

(a)

I I

EYE POSITION

I I I

HEAD ACCELERATION

I V

HORIZONTAL CAL. 1 SEC
VERTICAL CAL. 10 DEG.
& 10000 DEG/SECsq

CONTINUOUS CHANGE OF GAZE TYPES WITH CHANGE OF INTENT
(CHANGING HEAD-EYE LATENCY)

50 DEG

0 DEG

END OF HEAD MOVEMENT

BEGIN OF HEAD MOVEMENT

0.2 SEC

HEAD POSITION

50 DEG

0 DEG

TYPE: II    I    III    IV

50 DEG

0 DEG

EYE POSITION

-50 DEG    (b)
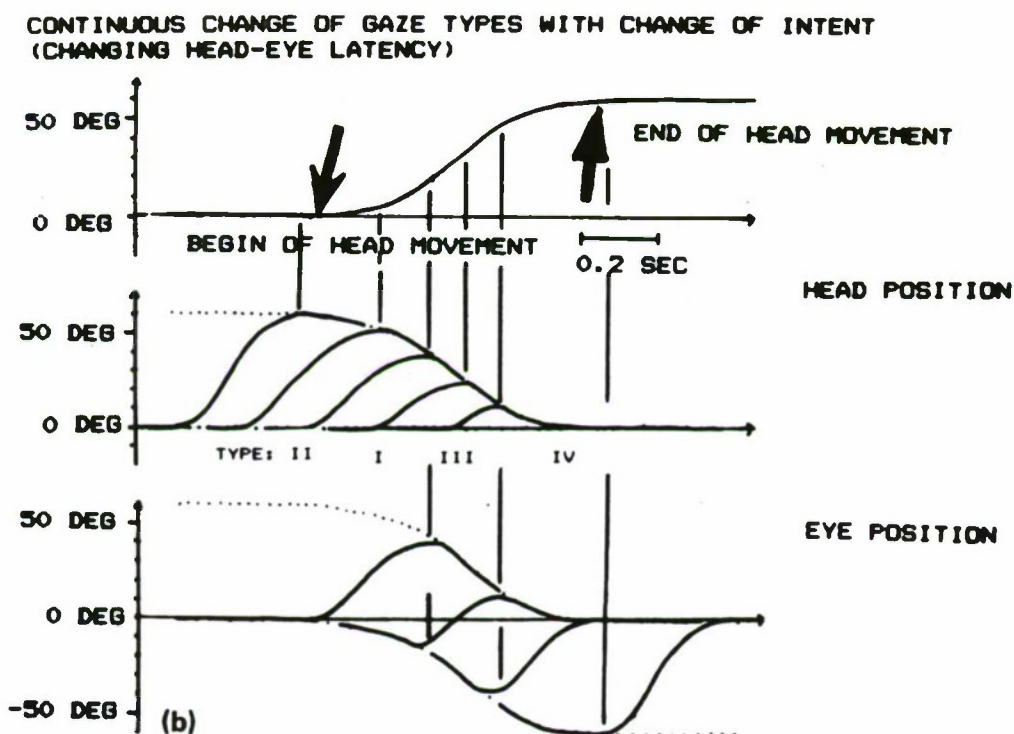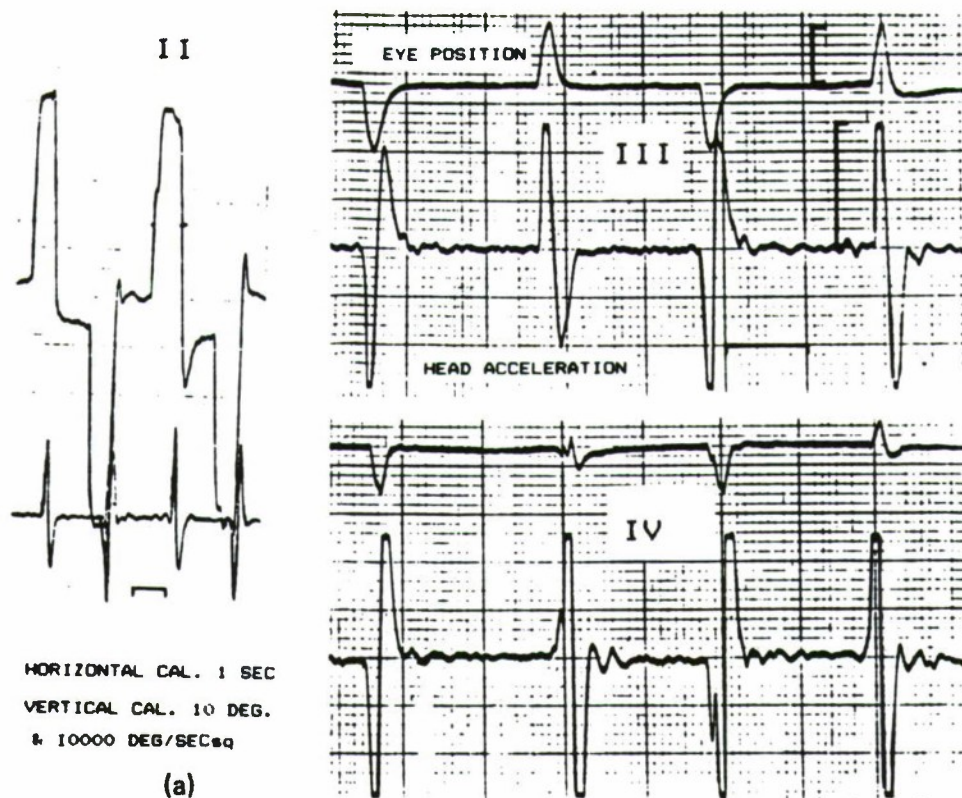
Figure 1.– a) Gaze types 2, 3, 4 generated intentionally through verbal feedback (upper).
b) Explanatory scheme for the continual change of gaze types (lower).

44-8

**(a)**

HEAD ACCELERATION (1000 DEG/SECsq = 1 BAR)

EYE POSITION (3 DEG = 1 BAR)

TIME CAL. 0.102 SEC

LARGER VOR RESPONSES TO RANDOM PERTURBATION

**(c)** WITH LOWER CONCENTRATION

EYE POSITION (3 DEG = 1 BAR)

**(d)** HEAD ACCELERATION (1000 DEG/SECsq)

SUPPRESSED AND DELAYED VOR RESPONSE
TO RANDOM PERTURBATION FROM
P,V,A = O WITH HIGH CONCENTRATION

**(b)** RANDOM PERTURBATION WITH P,V,A = O

WITH LOW AND HIGH CONCENTRATION

**(e)** PASSIVE ACCELERATION FROM ZERO VELOCITY

TIME

0.1 SEC

LOW INTENT

20 DEG

HEAD ACCELERATION (1000 DEG/SECsq)

INTERMEDIATE INTENT

15 DEG

EYE POSITION

MAXIMAL INTENT

10 DEG

Figure 2.– Random perturbation from primary position. (a) Low and (b) very high intent.
Random perturbation from primary position, (c) low and (d) very high intent. Random
perturbation from primary position, explanatory scheme (e).
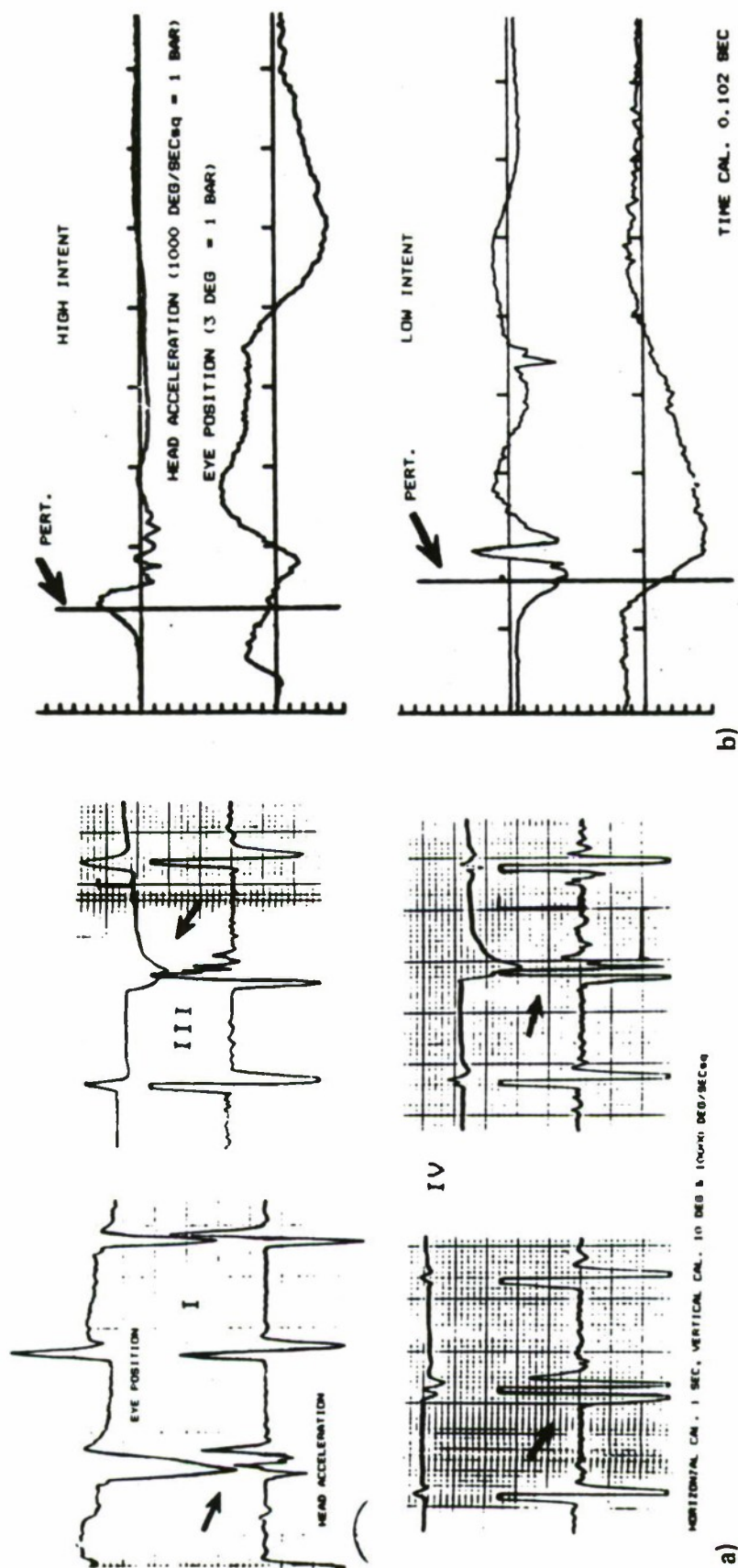
Figure 3.– a) Random perturbation of gaze type 1, 3, 4 in flight (arrows; left). b) Random perturbation of gaze type 3 with high (upper) and low (lower) intent (right).
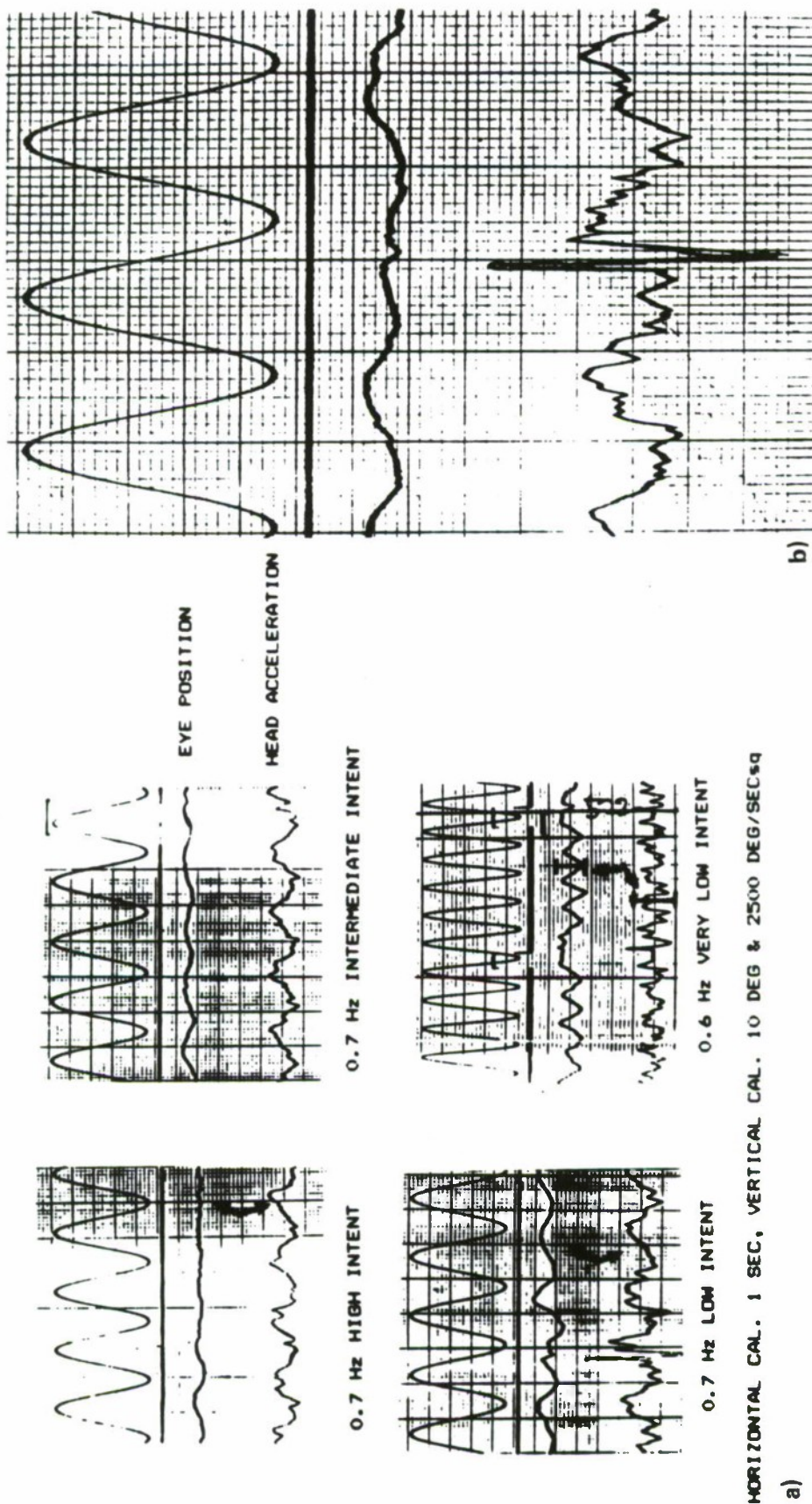
Figure 4.– a) Variable amount of fixation suppression of VOR as a function of intent (left).
b) Random perturbation of coordinated gaze pursuit: suppressed VOR response (middle).

# COMPUTER GRAPHICS

# THE MAKING OF THE MECHANICAL UNIVERSE

James Blinn
JPL Graphics Laboratory
Pasadena, California

## EDITORIAL FOREWORD

The *Mechanical Universe* is a two-semester, introductory level, television-based physics course. In the fall of 1985 the first semester of *The Mechanical Universe* was released to the academic community and public broadcasters. The two semesters of the course, *The Mechanical Universe* and *Beyond the Mechanical Universe,* consist of 26 half-hour television lessons and two versions of a text, one for science and engineering majors and the other for nonmajors. The course is scientifically sophisticated and mathematically rigorous, teaching and using calculus. The lecture programs contain computer animation used as a primary tool for the instruction in physics. Each program begins and ends with Caltech Professor David Goodstein providing philosophical, historical, and often humorous comments from his lectures at Caltech.

The television series is not only the basis for a college course, but it also is suitable for a general audience interested in stimulating and challenging science programming. The *Mechanical Universe* television series and college course were funded by The Annenberg/CPB Project and The National Science Foundation (Calfornia Institute of Technology, 1986).

The following sections excerpt a number of design considerations regarding the dynamic computer graphics used to communicate physical phenomena and mathematical principles included in the *Mechanical Universe* (Blinn, 1987). The specific recommendations were not intended to be freely extended to other graphics interface applications, but do represent the considered judgment of a pioneer of computer graphics and certainly identify design issues that are faced in all attempts to use computer graphics as a medium for communication of spatial information.

## CHAPTER 1 – OVERVIEW

### 1.1 INTRODUCTION

The *Mechanical Universe* project required the production of over 550 different animated scenes, totaling about 7 1/2 hours of screen time. The project required the use of a wide range of techniques and motivated the development of several different software packages. This report is a documentation of many aspects of the project, encompassing artistic/design issues, scientific simulations, software engineering, and video engineering.

My interest in *Mechanical Universe* is twofold. One, to produce the material and two, to see what tools need to be developed. It is hard to develop tools if you don't know what they are supposed to do. Having a large animation project provides a lot of experience on what the

problems really are, instead of what somebody thinks they might be. This is a somewhat empirical approach to systems design. That is, several special-case systems are built, motivated just by the needs of some particular project. They are then analyzed to see what things they seem to have in common. In doing this sort of examination, it is important to realize that you cannot prove that your assertions are correct in the same sense that you can prove a mathematical theorem. The best that can be said is that the mechanisms described here seem to work well for the problems to which they have been applied.

In this section I will discuss a few ideas on graphical design in general. The emphasis will be on concepts that are not specifically for scientific animation, but those that may be applied to other uses of visual communication.

I haven't learned this by formal training. It has come by practice, intuition, and perhaps genetics (I come from a family of artists). I learned to solve design problems by being presented with them and by being forced to think about the implications of color and shape choices. The results are what made sense to me at the time.

## CHAPTER 2 – GRAPHICAL DESIGN (STATIC)

Static design refers to the appearance of a single frame. The concept of motion design is discussed in the next chapter.

## 2.1 WHAT IS A DESIGN PROBLEM?

Let us begin with the question, "what is a design problem?" It can be likened to pantomime. You must present some information that, perhaps, could be described in words, but you are required to use only pictures.

Some examples:

- The Voyager spacecraft approaches a planet. A moon is off to the side. You must pan across to see it, but still give the viewers some idea of context of where they are now looking, compared with where they were looking before.

- How about a more detailed example? We will take an example from program 5, Vectors. The idea is to list the various types of vector expressions and to give an idea of whether the result is a vector or scalar. New items are added to the list as the program proceeds. The whole list may not fit entirely on the screen. In addition, as a new item is added, some geometric demonstration is needed to show what it is.

Let's look at a solution to this last example. We represent an abstract "space" where the vectors live as a kind of vector land. There is a river running down the middle separating it from scalar land. This allows us to display the lists in perspective receding into the distance. As each new object is introduced, it is added to the front of the list and the list recedes farther into the distance. Old list items may no longer be legible, but the memory of them is enough to

remind the viewer of what they are. The key elements are to (1) differentiate between vectors and scalars and (2) give an impression of three-dimensional (3-D) space, but not to make it look too realistic.

An oblique view of the ground plane must appear to recede into the distance. This can be shown by texture. An obvious texture is a grid which shows perspective very well. However, at this point in the academic development, the notion of a coordinate system has not yet been presented. Some other textural effect must be used. Texture mapping a random, say pebbly, texture would be slow. The resolution is to place a randomly scattered group of lines looking like grass across the plane. Just a few such lines can give a very cheap impression of receding ground plane. Also, the color of the plane is made to get bluer and paler as it moves into the distance.

Drop shadows help to bring out the 3-D quality and make the vectors seem to hover above the plane, giving an interesting surreal effect.

Later in the program, when unit vectors and coordinates are introduced, the grid is placed on the plane (but only a small piece of it). Grids are a bit overused in computer graphics, but for much of what we do at *Mechanical Universe,* they are necessary because we are actually plotting graphs.

When we introduce unit vectors $\hat{c}$ and $\hat{s}$, they tip their hats. When we show the construction of a vector product, the term for *vector add* and *vector multiply* are slid down close to the grid.

## 2.2 DIRECTION OF ATTENTION

It is necessary to direct the attention of the viewer to the important parts of the picture. Scenes are shown on television in fairly brief bursts, so the important parts must stand out. One good trick for doing this is to look away from the screen and look back quickly; determine what you see first when looking back. Is that the important part of the picture? If not, change the picture to make it so.

This means avoiding gaudy backgrounds; the background should not look more interesting than the foreground. In one example I had an equation over a dark blue background that graded into orange, giving a sort of sunset effect. It was very pretty, but the problem was that when you first looked at the screen, all you saw was the orange. I changed the background to a more neutral color and now the first thing you see is the equation.

## 2.3 AVOIDING INFORMATION OVERLOAD

I consciously avoid trying to "dazzle" the viewers. Dazzling implies an overload or numbing of the senses. The idea is to communicate and draw the viewers in instead of making them tip backwards off their chairs.

For the same reason, I don't use lots of spinning or tumbling of 3-D objects. It's distracting. There is a trade-off here between not giving your audience enough views of an object to be able to understand its 3-D shape versus making it confusing by spinning it around too quickly.

One important trick to encourage simplicity is to arrange for the designs to be done while viewing a monitor from across the room. If the image can be made legible at a distance of 10 ft, it's about right. This discourages putting in too much small detail.

## 2.4 COLOR SELECTION

Given the color television medium, we have both the opportunity to make scenes in color and the responsibility to make the colors look good. There are a few tricks to use in color selection.

I have favorite colors; I lean toward blues and greens. However, I don't like purple. I once used it purposely to break out of a rut, as a background in the scene on conic sections. I originally wanted to put a red cone in front of it, but I couldn't get a red that didn't disappear into the purple in dark areas (as seen in black and white). Finally, I went to a brighter yellow cone.

### 2.4.1 Make it Work in Black and White

When designing, look at the picture with the color turned off and see if it "reads" (to use a designer term). Reads in this context means "can you tell what is going on; do the appropriate things stand out?"

While color is important in the *Mechanical Universe* animations, it is not the only thing that differentiates items on the screen. It's not crucial. I have made consistent color decisions, but the viewer is not expected to remember color schemes to understand a scene.

### 2.4.2 Context

Color selection programs are minimally useful because colors always look different in context. The only real way to see how they look is to make an actual picture of the scene.

### 2.4.3 Distance Cues

Distance can be represented by making things disappear into a fog. This was done literally in a scene of the molecular arrangement of a salt crystal.

Other color cues: the color of things gets bluer and paler with distance.

Field lines are a complex set of 3-D curves. They can look like a pile of spaghetti if you're not careful. The distance effect is aided by three things: (1) normal depth cueing (things get darker—i.e., less luminance contrast—with distance); (2) drawing them in depth order so a closer (brighter) line will overlay a farther line; (3) making the intensity of the line darker at the edges than in the middle. This gives a slight "cylindrical" solid quality to the lines.

### 2.4.4 Not Too Many

Don't use too many colors.

There is a problem with running out of colors. There are more physical quantities to represent than there are easily distinguishable colors. You can't use saturation or value to distinguish things because sometimes these need to be adjusted depending on context, e.g., energy.

### 2.4.5 Consistency

Consistently use color schemes to recall previous results as well as to differentiate things. We will discuss the color scheme later:

- But the color scheme wasn't always consistent
- Paler colors for mass multiplied by something
- Colored backgrounds for two integrations of gravity law
- Colored backgrounds for bringing external equations to prove Kepler's third law
- Blue texture for energy equation

### 2.5 2-D/3-D CONSIDERATIONS

Two-dimensional diagrams are easier to understand than 3-D, especially when they are in motion. This is partly because labels keep getting in the way of 3-D diagrams in some views. Most of the physics of the first term of *Mechanical Universe* is essentially 2-D problems (like Keplerian orbits). These remain 2-D. The inherently 3-D concepts are torque and angular momentum. The punch line is, use 3-D only when absolutely needed.

In fact, some 3-D situations were simplified to 2-D. For example, I used 2-D for the Lennard-Jones atomic motion simulation and the ideal gas simulation. The actual physics is 3-D, of course, but 2-D shows the phenomena adequately and 3-D would be really confusing.

In the second term there were more inherently 3-D problems. You must use 3-D for electromagnetic fields. Many textbooks use 2-D for fields, but much is lost.

Three dimensions are also used as a trick to put more text on the screen. As the screen tilts back, more text fits. The top row might not remain legible, but we can remember what it was.

## 2.6 MAKING THINGS STAND OUT FROM THE BACKGROUND

Drop shadows help make things stand out from the background. While they are good for labels on graphs, don't put a drop shadow on the plotted graph line because it detaches it from the grid.

Put 3-D shadows for 3-D vectors even if there are abstract shapes with no light source. One can more easily see a 3-D shape by simultaneously having two views of the object, a 3-D view and a projection of that view on the xy plane. This is what the cubists were trying to do—show many views of an object at once. The shadow technique is more the way we are used to seeing and interpreting things.

Make the background a different value; use pale colors.

## 2.7 REALISM VERSUS ABSTRACTION

Images representing some real, physical object are often overlaid with labels, vectors, etc. For such scenes, the real object is rendered with a simulated light source and shading (usually with a simple polygon rendering program). The mathematical abstractions are overlaid with a line drawing program (lines don't change thickness as they get closer or farther from viewer).

## CHAPTER 3 – GRAPHICAL DESIGN (DYNAMIC)

From reading Thomas and Johnson's book (1981), you are left with the impression that animation is the highest form of human art. It encompasses all aspects of static art and adds timing and motion, too. Motion design may well be the next great research topic in computer graphics. Results shown here are very preliminary.

## 3.1 INTERPOLATION

It is the popular wisdom in animation that spline interpolation is better than linear interpolation. It is smoother. Most of the animations were done with splined motion. However, later in the series I began experimenting with linear interpolation and found it quite pleasing. Let's face it, the algebraic motions represent mechanical operations, so why not make them mechanical looking? In this case non-natural (jerky) motion sometimes looks more interesting than smooth motion because it's different and contains more high frequencies at the key frames.

## 3.2 INCORPORATION OF "CLASSIC" TECHNIQUES

There are various classic techniques that are found in "conventional" animation that apply here.

### 3.2.1 Squash/Stretch

Squash and stretch refer to a distortion applied to the shape of an object when it undergoes acceleration. This is easily done by animating the x and y scale factor of an object. Before it begins to move, it gathers itself up by shrinking in x, then it stretches out in x as it is moving, and when it stops it shrinks briefly and returns to its normal size. This wasn't done in the *Mechanical Universe* as much as it should have been.

### 3.2.2 Overlapped Motion

The concept of overlapped motion states that motion 2 should start before motion 1 is completed. This works well with character animation, but I found it of limited use in algebraic animation. In algebra there is just too much to follow as it is, without having the individual steps of a derivation merge into each other. Making the steps disjoint in time gives the viewer a chance to absorb one step before another begins. I did make the x and y motion of an object overlap, but this just rounds off the corners of the motion.

## 3.3 PERCEPTIONS OF SPEED

I found it interesting to discover how limited our perception of velocity is. Given two successive scenes, where an object moves, say, one and a half times as fast in the second scene, it is very hard to tell which is which. This was proven because we were showing velocity changes in a lot of the physics. Most of the solutions involved representing velocity spatially as well as temporally by adding streaks or velocity vectors to moving objects.

Another interesting speed-perception discovery concerns double framing. One would think that all animation is ideally single framed. Double framing is just an economy measure if you don't have the computer time to do all the frames. Double framing looks jerkier. But there's another perceptual effect of double framing—double-framed motion looks faster than single-framed motion.

That is, if an object moves across the screen in 1 sec, it will look like it is moving faster if it is animated as 15 frames double-framed rather than 30 frames single-framed. This was alluded to in Thomas and Johnson's book (1981) on Disney animation. They said that motion was sometimes purposely double-framed to give it a "jaunty" look.

### 3.3.1 Audience

When doing something of this nature, it is important to keep the audience in mind. I had a very specific audience in mind when I designed these animations before I understood the concepts. For the most part, these are the explanations that I would have liked to have had, and that would have made the most sense to me when I was learning physics.

### 3.3.2 Roots

We are all products of our environment. I would like to mention some previous experiences that have affected my design motions here.

Lillian Lieber and Hugh Lieber are a mathematician/artist team that produced a series of charming books in the 1940s. Hugh, the artist, had a very surreal sense of making mathematical symbology visually interesting.

Various Disney animations were produced for science and mathematics. Among these were "Man in Space" and "Donald Duck in Mathemagic Land."

George Gamow (1967) wrote several books popularizing physics. His best creation is the Mr. Thompkins series. In these books, Mr. Thompkins attends a physics lecture and falls asleep. In his dreams the physical point of the lecture is illustrated, usually by exaggerating the effects so they were more noticeable in daily life. Particularly memorable was a scene in the "Old Woodcarver's Shop" where a sculptor makes atoms out of little green marbles (electrons) and little red marbles (protons).

The "Chem-studies" series of films were made for high school use. These had several conventionally done animations of molecular dynamics during chemical reactions. The motion of the atoms in these animations beautifully gives a sense of the energetics of atomic bonding. These were produced by David Ridgeway, who is on the national advisory committee to the *Mechanical Universe*.

The Bell Labs produced science films such as *The Unchained Goddess* and *Our Mr. Sun*. These were directed by Frank Capra, a Caltech graduate, and also a member of the advisory committee to the *Mechanical Universe*.

Finally, a telecourse from the past: "Continental Classroom"; this was a for-credit course offered on television in about 1960. It had classes in mathematics, physics, and chemistry. When I was young I was interested in this stuff, but I didn't know where to go for information. When I found this course I got up religiously each morning at 6 a.m. to watch it. I understood only about half of it, but it kept my interest in the subject alive. I hope that, with the *Mechanical Universe*, I might be making a series that generates similar interest in a new generation of students.

# CHAPTER 4 – VISUAL METAPHORS (DESIGN)

In this chapter I will discuss visual metaphors for physics, grouped by design concepts.

## 4.1 COLOR

A normal textbook diagram has shapes, lines, and text. In video we have, in addition, color and motion. The challenge is using them. Motion usage is, for the most part, more obvious than color usage. Where there is some previous convention for color assignment, I tried to use it. Where there was none, I had to invent one.

When referring to explicit color values, I will use the notation developed by Alvy Smith. Color is three numbers representing

1. Value or brightness (0...1).

2. Hue going around the color wheel. Numerical quantities go from 0-5 for one cycle: 0 = red, 1 = yellow, 2 = green, 3 = cyan, 4 = blue, 5 = magenta.

3. Saturation. 0 = neutral, 1 = fully saturated.

Written, as an expression, (i,j,k) (i.e., (1,0,1)) would be a red of maximum brightness and saturation.

Many different ideas were keyed to colors. Much of this was subtle, and the animations never relied solely on the color to be understandable. I was left with the impression, however, that there simply aren't enough colors to have a unique one for *everything*.

### 4.1.1 For Dimensional Analysis

When physical abstractions such as acceleration or torque are represented in vector diagrams or algebraic labels, there must be some color. Rather than just making all vectors and labels white, I chose to institute a color scheme that is keyed to the units in which the quantity is measured. These color schemes are maintained throughout the series. This provides for a sense of continuity and also gives the viewer a sense for dimensional analysis.

Also, I tried to avoid the temptation to get overly cute with the colors. Colors are used primarily for labels. Terms in equations are usually white; otherwise, the equation tends to look like confetti. A term is shown in color only if the dimensions are important for a particular derivation.

Position, velocity, and acceleration are the most commonly used quantities. Position was a green = (1, 1.8, 1); velocity was a yellow = (0.7, 1.2, 1); and acceleration was a red = (1, 0.2, 1).

There are several motivations for this general color scheme. As successive derivatives are taken, the color shows a smooth progression along the color wheel from green to red so there is a visual progression between the colors. (Actually, the reddening applies not so much to derivatives as to the division by time.)

Acceleration is the most "active" of the three concepts. But red means "stop," not a very dynamic idea (although it takes deceleration to stop). This might be a counter argument for the use of this color. But red is also the most exciting, attention-getting color. It shows that something is going on, and thus looks dynamic.

Green (as in grass) shows a static "place-like" effect.

This color scheme worked well when applied to a scene showing an abstract bicycle rider. The intent was to show elevation and slope. The normal color for informational traffic signs (green) was used to label the elevation. The normal color for warning traffic signs (yellow) then labeled the slope.

Note that the colors chosen are not pure; the hue values are not integers. The exact hues were selected visually to look nice together. Exact primary colors tend to look boring.

Mass times acceleration gives force. Mass times velocity gives momentum. Force and momentum were given the same colors as acceleration and velocity except that the saturation was reduced. I think of mass as a sort of dark grey color, looking solid, like lead or iron. So adding grey to the above colors desaturates them.

Energy is a dark blue color. This was chosen to look sort of like a lightning bolt. Energy's color is (0.2, 4, 1).

Angular momentum is a sort of rotational concept. I toyed with the idea of giving angular momentum vectors a sort of barber-pole effect, but it seemed too busy. Angular momentum is also mass times velocity times distance. Maybe a sort of pale yellowish-green? But that would not make it distinguishable enough from the other two. Finally, I decided to take off in a new direction and make it a pale blue. Torque, the derivative of angular momentum, is lavender (blue with red added to it).

Area and volume were made variants on the green color. Area is a slightly bluer shade. Volume is a still bluer shade. Maybe I was getting too subtle here, but you have to pick *some* color, and it might as well be for *some* reason.

Actually this choice was not entirely conscious, and as a result, the color for area is not exactly consistent through the entire series. For example, the color of Gaussian surfaces in the electricity programs was the position color, not the area color. This led to some problems when showing surface integrals. You do your best, but sometimes mistakes creep in.

## 4.1.2 Solid, Liquid, Gas

In the thermodynamics discussion there is a section on the states of matter. In particular, a PVT diagram is separated into regions where a substance is a solid, a liquid, and a gas. These regions were colored as follows.

Solid – medium brown; an earth color, designates the solidity of ground.

Liquid – bluish; like the color of water.

Gas – white; a transparent color.

In the PVT diagram there is a region above the critical point where the distinction between liquid and gas disappears. Van der Waals' equation was used to find the degree of liquidity and to calculate a saturation value smoothly grading from blue to white for this region.

## 4.1.3 Electric Charge

Positive and negative charges are shown in many scenes. There has been a sort of convention for some time in engineering to make the positive leads red. In addition, the books by George Gamow represented electrons as green marbles. So a similar color scheme was chosen for the *Mechanical Universe*.

But there are two problems here. First, not everyone has a color television set. So the colors were chosen so that, in black and white, they would still have enough difference in brightness to be distinguishable. Second, although red and green are complementary colors visually, in video it is red and cyan (a sort of pale blue). In some instances a neutral charge (e.g., for neutrons) is shown as, obviously, white. It would seem best to make the plus and minus colors add up to white. So a more bluish hue was chosen for negative charge. The exact value was actually changed during the second half of the series to be exactly cyan. This seemed necessary to make plus and minus add up to neutral, but I'm not sure it was a good idea in retrospect.

## 4.1.4 Electric and Magnetic Fields

I've always thought of magnetic fields as blue, and many published diagrams have shown it as blue. In fact, in an earlier project showing the magnetic field of Jupiter, I made the field lines blue. The question is, what color are electric fields? Since they are lines between positive (red) and negative (greenish blue), I decided to make it the color halfway between them, yellow. Note again that this is a different yellow than is used for velocity.

## 4.1.5 Relativity Coordinate Systems

There were many scenes in the relativity section that illustrated events as seen from two different reference frames. The two frames were usually those of a cartoon Albert Einstein and a cartoon Henry Lorentz. When they first appear, Albert is wearing a tan suit and Henry is

wearing a blue suit. Thereafter, any algebraic or pictorial reference to Albert's frame is drawn in tan and any reference to Henry's frame is blue. These colors were initially selected as typical colors that suits come in, but they were fine-tuned to show up distinctly in black and white and when placed on a common background. Actually, when I first decided to do this, I had made Henry's suit dark grey. But dark grey didn't look good as a comparison color to tan—tan and blue are more balanced complementary colors. I had to remake one of the first animations just to change the color of Henry's suit. The production people probably thought I was nuts.

### 4.1.6 Wave/Particle Duality

The last three programs of the series begin to touch on quantum mechanics. Several of the scenes depicted wave-particle duality. Complementary background colors were selected to represent particles and waves. All particle equations appeared over dark pale green; all wave equations and plots of wave functions appeared with a dark pale magenta background.

## 4.2 LITERAL VERSUS SCHEMATIC

My tendency is to be too literal. The sizes and timings of some phenomena sometimes have too big a range to make this easy. But, because this is computer animation, the viewer expects precision and accuracy. When sizes or timings must be distorted into schematic diagrams, it is important to give some visual cues that this is being done. One way to do this is to have the schematic scenes drawn with sketchy or irregular lines. This removes the precision effect of perfect lines.

### 4.2.1 Literal

Some things were done geometrically correctly, even though it was difficult. For example, the radii of the orbits of the Bohr atom are proportional to the perfect squares (1, 4, 9, 16, ...). To see as many as four orbits, the scale must be too small to make the first orbit clear. This was usually solved by having the camera pull back when discussing the larger and larger orbits. This is a useful general principle, as it was described in an earlier chapter concerning a list receding in perspective. If some things are too small, start close up and pull back.

### 4.2.2 Schematic

When force laws are introduced, we needed to show the operation of gravitational and electric forces. At this point, the magnitudes weren't important, only the signs. Crude schematic faces were used as mass particles (grey faces) and as positive and negative charges (red and cyan faces). The motion was sketchy, showing only attraction versus repulsion, and the faces were sketchy, with irregular and comical lines. This visual signaling was not done enough in the series.

Other scenes with schematized motion included:

- A depiction of resistance in metals. The normal velocity of electrons in a metal is far greater than the drift velocity, which is the electric current. Therefore, an accurate depiction of current wouldn't look much different than random thermal motion. The relative velocities were made more equal for illustration purposes. Also, resistance is caused by collisions of electrons with imperfections and thermal motions of the atoms in the metal lattice. These are usually too few and far between to be easily noticeable. They were made more obvious by flagging some metal atoms a different color and having the electrons bounce off them elastically, while not being affected by the positions of all of the nonflagged atoms.

- An electrical spark is generated by a chain reaction. Electrons are accelerated by an electric field and build up enough kinetic energy to knock other electrons off atoms. Again the typical spacing and frequency of the real situation would not fit on the screen. Some exaggeration was done.

## CHAPTER 5 – VISUAL METAPHORS (PHYSICS)

Here are some more visual metaphors, this time grouped by subject matter, rather than by design issues.

## 5.1 ALGEBRAIC BALLET

To make the science respectable we had a lot of algebra to present. Algebra, however, can be a bit draggy. We decided to liven it up by animating the algebraic transformations that the equations go through. These animations usually go by quickly. In fact, it is unlikely that the viewer will be able to follow all the steps upon first viewing. The speed was a concern, but we felt that making it slower would slow down the programs too much. The idea is to get the feel for what is going on and be able to look at a videotape slower to get the detail later if desired.

Transforming algebraic operation into motion proved to be an interesting exercise. Many of the motions seemed pretty obvious to me, but they are listed here for completeness.

### 5.1.1 Term Labeling

It's easy to lose track of what different symbols in an equation represent. This was addressed by having the symbols identify themselves with English words popping out and shrinking back into them.

### 5.1.2 Balancing Act

Simple algebraic operations to move terms around were animated literally.

- Terms moving to the opposite side of the = sign. Adding on one side means subtracting on the other so a + or - sign flips its identity as the term hops over the =.

- Factors moving to the opposite side of the = sign. Multiplying on one side means dividing on the other. When a factor jumps over the =, it lands below or above a division bar according to whether it came from above or below.

- Distribution: $a(b + c)$ becomes $ab + ac$ by having the $a$ jump up, split in two, and each copy land next to the appropriate term.

- Squaring: Either two 2s come down from above and land on each side of the =, or a 2 on one side of an = sails over and changes to a $\sqrt{}$ sign on the other side.

### 5.1.3 Canceling

This applies to the removal of identities like $a - a$ or $a/a$. Some ways used to depict this were:

- A lightning bolt zaps the two terms and they disappear.

- An eraser appears and erases the terms.

- The two terms turn red and fall off the bottom of the screen together.

- A video-game-style spaceship flies in and fires a missile to explode the term.

- A Monty Python-style foot stomps out the terms.

- The Hand of God touches the term and it becomes a puff of smoke. This was used in the program that derived Kepler's first law (orbits are ellipses) from Newton's laws. The program made comparisons between the accomplishments of mathematics and physics and the accomplishments of art, drama, and music. Art was represented by the Sistine Chapel of Michaelangelo with the Hand of God giving life to Adam. The essential cancellation in the math that makes the derivation work is $r^2/r^2$; this is done by the Hand of God, too.

- Multiplication sign snipping out a term: The expression $v \times v$ is equal to 1. When this appears, the cross product sign magnifies around the surrounding v's and then squashes rapidly in $y$, snipping out the terms.

- Simply fading the terms out: This, of course, was the simplest and was done the most often.

### 5.1.4 Recalling Old Results

When a result from a previous program, or from a previous course is introduced, some effort was made to indicate to the viewer where it came from. Some examples are:

- A trigonometry book flies in, opens, and trig identities fly out.

- A head with a hinged lid opens to receive some intermediate results; later it returns and the intermediate results fly out.

- A hand pulls down a window shade with old energy equations.

- An entire scene is reprised from a previous program.

- Some results were derived against a background image of some distinctive color. Later, when the results are needed, a slide comes in containing the equation with the same background as old scene.

### 5.1.5 Substitution

Substitution involves taking an equation defining some variable and replacing occurrences of that variable into another equation. Some examples:

- Vertical shrinking. A term is replaced with a number by shrinking the term vertically to zero and having the number expand up from zero in its place.

- Vacuum cleaner. The identity equation appears above the main equation. The replaced term from the lower equation moves up to the identity to merge with its copy there. The other side of the identity equation moves down to the empty spot left in the original equation.

- Several calculus identities (such as turning $dr/dt$ into $v$) were shown by rotating the $dr/dt$ about the $y$ axis and having it become $v$ when the other side appeared.

### 5.1.6 Jokes

The program on wave motion shows some approximate relations between wave speed and various physical parameters. The $\approx$ sign ripples like a propagating sine wave while these equations appear. This was done by modeling the lines of the $\approx$ sign with a one-cycle helix. Rotating it about $x$ and then scaling by 0 in $z$ made it ripple.

## 5.1.7 Calculus

A few algebraic operations on calculus notation:

- $d/dt$ flies in from left and impacts $f$ to form $df/dt$.

- The $f$ slides up and down to form $(d/dt)\,f$ from $df/dt$.

- The two symbols $\int$ and $dt$ move in on either side of $f$ and clamp it together to form $\int f(x)dt$.

- A simple differential equation like $(dx/dt) = y$ is solved by moving $dt$ to the other side to make $dx = y\,dt$. Then the left-hand $d$ hops over the equal sign and changes into a $\int$ sign, to make $x = \int y\,dt$.

- Integration is done by the $\int$ sign ratcheting across an expression, sort of like a credit card imprinter.

- $\oint$ is formed by drawing the circle on the $\int$ as the path of integration is traced out in a geometric diagram in the background.

- $\oiint$ is formed by revealing the circle on $\iint$ as a Gaussian surface is spread out around a volume in a parallel diagram.

## 5.2 CALCULUS

### 5.2.1 Limits

Use explosion to express the limiting process when $\Delta$ turns into $d$. The explosion was generated by a simple 2-D pattern scaled up and faded out simultaneously.

### 5.2.2 Symbolic Derivative Machine

Because we evaluate derivatives and integrals symbolically many times in the series, we developed a quick way to do it—the derivative machine.

**5.2.2.1 Design**– The derivative machine is an expression transformer. It has two functions—differentiation and integration. An expression goes in one end and comes out the other end, so it needed to be thin in the $x$ direction so there would be plenty of room on each side to show the inputs/outputs. When the derivative machine is first introduced, it comes in a crate marked "ACME Derivative Machine" (a hat tip to the old Chuck Jones Roadrunner movies). A crowbar shaped like an integral sign opens the crate.

Some random wheels and lights made it look Rube Goldbergish. The sides are not exactly straight and the wheels are not exactly round.

**5.2.2.2 Internals–** When the derivative machine is introduced in program 3, the internals are shown two ways:

1. As various elementary operations are introduced, they shrink down into a sort of circuit board that is plugged into the machine, the door slams, and a new light blinks on on the front panel.

2. An alternative view of the internals was given briefly, showing the details of how the elementary operations are applied to take the derivative of the simple expression $x^2$. This was intended to be somewhat a metaphor on how symbolic derivative computer programs work. The input function comes in on a conveyor belt. An eyeball on a stalk comes down and looks at it. (This is indicated by a dotted line running from the eyeball to the function.) This is the pattern recognizer. The derivative operation is basically one of matching the desired function against a list of known patterns which are pulled down into the scene like window shades. Then the proper pattern is found and checked. There will be some dummy parameters in the pattern which need to be filled in with the specific terms from the equation. The eyeball observes these and some handles come down and simultaneously turn all occurrences of the dummy parameter into the specific term needed. Identities such as $x + 0$ or $x * 1$ are removed by an eraser. The expression $x + x$ is turned into $2x$ by a vise-like adder. The final expression is carried out on a conveyor belt.

**5.2.2.3 Operation–** The lever on the top controls the operation of the derivative machine. When you throw the lever to the right, it takes an expression in the left hopper and spits the derivative out the right hopper. When you throw the lever to the left it takes an expression in the right hopper and spits out the antiderivative (integral) on the left. Sometimes the expression stays put and the derivative machine passes over it. Note: it doesn't evaluate integral expressions, it just takes the antiderivative (i.e., you don't feed $\int x^2$ in to get $(1/3)x^3$, you just feed in $x^2$). As it operates, the horizontal and vertical scales cycle up and down a bit to give it a squash and stretch look.

## REFERENCES

Blinn, James F. (1987) The mechanical universe: an integrated view of a large animation project. (Course Notes: Course #6) Proc. 14th Ann. Conf. on Computer Graphics and Interactive Techniques, ACM SIGGRAPH and IEEE Technical Committee on Computer Graphics. Anaheim, Calif.

California Institute of Technology (1986) The Mechanical Universe ... and beyond. Corporation for Community College Television. Pasadena, Calif.

Gamow, George (1967) Mr. Thompkins (paperback) Cambridge University Press, Cambridge.

Thomas, Frank and Johnston, Ollie (1981) Disney animation: the illusion of life. Abberville Press, New York.
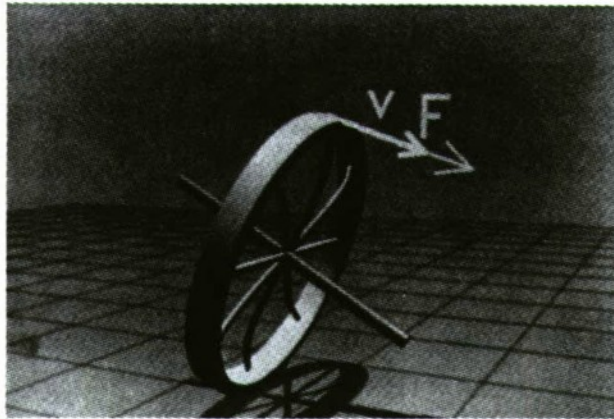
Computer animation dissects the forces and motions that make a gyroscope do its tricks.
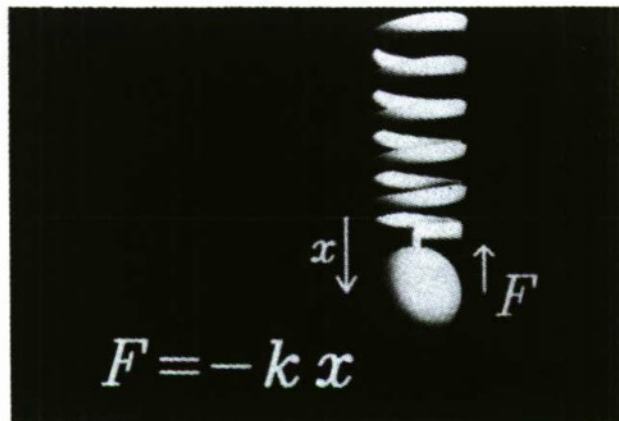


The spring force, or Hooke's law, is described in this animated scene from the Harmonic Motion episode.
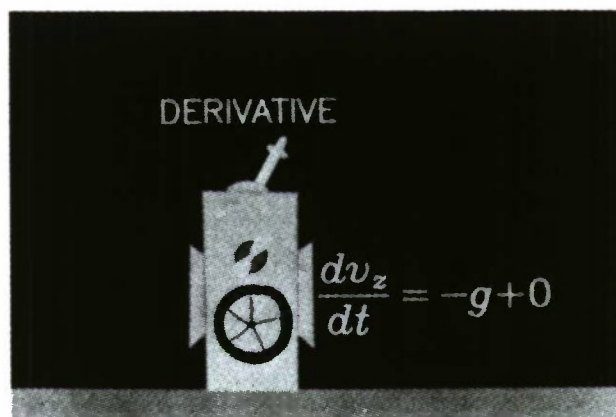


The *Mechanical Universe* derivative machine has become a legend in its own time.

# SYNESTHETIC ART THROUGH 3-D PROJECTION: THE REQUIREMENTS OF A COMPUTER-BASED SUPERMEDIUM

Robert Mallary
ARSTECNICA: Center for Art and Technology
University of Massachusetts/Amherst
Amherst, Massachusetts

## SUMMARY

A computer-based form of multimedia art is proposed that uses the computer to fuse aspects of painting, sculpture, dance, music, film, and other media into a one-to-one synesthesia of image and sound for spatially synchronous three-dimensional (3-D) projection. Called synesthetic art, this conversion of many varied media into an aesthetically unitary experience determines the character and requirements of the system and its software. During the start-up phase, computer stereographic systems are suitable for software development. Eventually, a new type of illusory-projective "supermedium" will be required to achieve the needed combination of large-format projection and convincing "real-life" presence, and to handle the vast amount of 3-D visual and acoustic information required. The influence of the concept on the author's research and creative work is illustrated through two examples.

## INTRODUCTION

The concept of synesthetic art described here is the product of an approach to art that looks to science and technology for the invention of new media for art, and to new media as a way of expanding the aesthetic, stylistic, and expressive possibilities of art. That science and technology indeed have the capacity to play this role was demonstrated in the last century by the invention of photography and cinematography, and more recently by the invention of television. That not every application of science and technology to the visual arts has this impact, however, is demonstrated by the history of kinetic sculpture and other kinds of technologically oriented art that have appeared over the last 40 years, none of which have acquired the importance of these earlier inventions or developed into an authentic and accepted new art form (ref. 1).

In 1967, on learning that the computer, in addition to everything else it can do, is able to generate and process images, I asked myself whether this amazing brain-like technology would eventually provide the basis for a new form of art comparable in importance to photography and film. On deciding that the computer indeed has this potential, my next question concerned the character of this new form of art and the role of the computer in its production. While these ruminations took place without benefit of such terms as synesthetic art or supermedium, the concept I developed, though somewhat vague compared to my way of thinking about it now, was essentially the same as the one proposed and described here (ref. 2).

Before providing a systematic outline of synesthetic art and its requirements, it may be helpful if I briefly describe what I mean by synesthetic art and what I visualize when using the

expression. The best way to do this vividly and expeditiously calls for a small exercise in "imagineering."

Think of an empty transparent block of space about the size of a 19-in. computer-graphic color monitor, its depth the same as its height. Fill this space with a collection of floating objects that vary considerably in shape and size, some small and spherical like marbles, others larger and more irregular in shape. Then add something quite different to the mix, something like a luminous cloud or foggy mist. Endow this combination of solid forms and vaporous intangibles with colors, textures, patterns, shadows, and other attractive qualities and attributes.

At this point, set the ensemble in motion, into a choreography of disappearing and reappearing, swelling and contracting, disintegrating and reassembling, changing one into another and back, and into arrays of identical objects that move choreographically to the distinctive sounds of computer music. And note that the sounds are fully as spatialized as the visual material, with many of them moving in precise spatial synchrony with them.

Though the dominant effect is more abstract than realistic, there are hints of the real world here and there. Whether abstract, realistic, or something in between, the objects pass eerily through one another, completely unhindered by visible mechanical or electrical assistance. Aspects of painting, sculpture, photography, cinema, and dance fuse into an ambience of near transparency, with objects apparently farthest from the eye nearly as visible as those that are near. With forms melting into air and air into forms, the overall effect, despite the prevailing three-dimensionality, is as pictorial–even as "painterly"–as it is sculptural. And because of the patterned and formalized movement, the affinity with choreography and the dance is as obvious as the connection with painting and sculpture.

These imaginary events in an imaginary block of space are as far from the synesthetic art of the future as they are from any method of three-dimensional (3-D) projection available today. Yet, with only minimal trouble and expense, the color monitor of an Atari 1040 ST personal computer can be converted into a not-too-crude approximation of the imaginary block through the purchase of a set of liquid crystal stereo goggles (ref. 3). The Atari is low resolution. A more expensive stereographic system with higher resolution, however, if adapted to a large-format, video projection system, could expand the block and the events within it to a scale of 6 x 8 ft or more (ref. 4). Eventually, if my confidence in the future of synesthetic art is justified, the scale of the block will be measured in yards rather than feet; the quality of "reach out and touch it" realism will be overwhelmingly convincing, and the varied happenings within the huge block of space will be correspondingly impressive (ref. 5). The computer-based method of 3-D projection that can achieve this near-perfect realism on such a scale is what is meant by the "supermedium" mentioned in the title. Though it is not impossible that this supermedium will emerge as an outgrowth of the stereoscopy and holography we know today, the limitations of both are just as likely to prove insurmountable.

In order to stress that synesthetic art is as much concerned with sound as it is with pictorial and sculptural kinetics (and eventually, with drama, performance, and narrative content as well), the block of spatial activity will henceforth be referred to as an "event space." It could just as well, however, be called a "stereo event space," in acknowledgment that 3-D projection by computer stereographics, despite its limitations, will probably incubate development of synesthetic art for many years to come.

# GENERAL FEATURES OF SYNESTHETIC ART

Synesthetic art has four essential features that determine the design and operation of the computer-based system and how it is used to create synesthetic art. These features refer to (1) the comprehensive multimedia character of synesthetic art, a feature that calls on the system to either capture or simulate a wide variety of attributes and materials, both visual and acoustic, drawn from many different forms of art; (2) the bimodal spatial synesthesia of image and sound, a feature that enables the system to superimpose visual and acoustic elements and move them together in the illusory-projective event space; (3) the aesthetically integrated character of synesthetic art, a feature that calls on the system to assist in organizing these disparate materials into a close-knit synesthetic unity (an option, not a requirement imposed on users of the system); and (4) the extremely integrated and task-oriented character of the system itself, a feature that calls on the developers of the system and its software to take full advantage of the computer's ability to capture, generate, process, and spatially manipulate both images and sounds by drawing upon resources as diverse as computer graphics, image processing, computer music, and artificial intelligence, among the many germane fields and disciplines.

The block diagram in figure 1 represents all four of these features in a general way. More concretely, however, it also represents the five major blocks of software comprising the entire synesthetic package of programs, along with the quite specific requirements associated with each of these blocks. The diagram conforms to the standard format for such graphic representations, with input at the top, output at the bottom, and everything associated with the ongoing manipulation and control of the total mass of visual and acoustic information presented in the large central panel, coded in light grey.

# REQUIREMENTS OF THE SYNESTHETIC SYSTEM

Some of the more specific features of synesthetic art can be gleaned from this summary of the requirements imposed on the computer system, because many of the features, functions, and requirements of the system provide mirror reflections of synesthetic art itself.

## Realism

During the start-up stage of synesthetic art, a high degree of realism is hardly an achievable, or even desirable, objective. From the very beginning, however, some use of low-resolution, generalized forms of realism are necessary, first for aesthetic variety and interest, and second as steps in the direction of the narrative and dramatic realism associated with the long-term full multimedia potential of synesthetic art. As an aspect of this eventual development, the degree of near-perfect realism should be such that an observer peering casually into the event space might easily fail to distinguish between the projected image of an object and the actual object itself. This can be thought of as the ultimate "Turing test" for 3-D projection, a level of "reach-out-and-touch-it" realism that may never be fully achieved, but that is useful nonetheless as an unambiguous standard and objective for ongoing research and invention (ref. 6).

Realism in synesthetic art, in whatever degrees and varieties it appears, will be achieved through a method of real-world image capture that is basically photographic, whether involving video, cinematography, or some other method. Or it will be achieved within the computer itself through image synthesis "from scratch," using mathematical and algorithmic techniques along the lines of solids modeling and ray tracing. Or the realism will be achieved through combination of both of the preceding, or perhaps through a method yet to be developed. A key requirement would seem to be a method of 3-D capture that digitizes the information as it is acquired, facilitating its transmission into the computer and submission to the myriad form transformation operations basic to this concept of synesthetic art.

## Abstraction

The second requirement shifts away from realism to the opposite end of the stylistic spectrum in demanding that the system supply an endless variety of visual qualities and attributes having as much to do with abstract art as with realism. These attributes, which are at the core of synesthetic software along with objects they enhance, pertain to such basic elements as form, shape, color, texture, pattern, tone, translucency, hard and soft edges, optical distortions, etc. Ideally, any of the styles and iconography associated with 20th century visual art and its media–starting with painting and sculpture, but also including photography, printmaking, computer graphics, computer animation, video art, abstract film, laser sculpture, and light art–should be capable of being simulated and, if necessary, translated into an effective 3-D equivalent idiom for integration into the synesthetic mix. In time as the synesthetic software package expands, synesthetic artists should be able to work in virtually any style conceivable, with no constraints other than those self-imposed for expressive or aesthetic reasons. The objects mentioned in the Introduction, and their mutations as arrays, regions, and total event spaces, are represented in the block diagram under the general heading of "visual/spatial components."

## A Choreography of Change and Motion

The third requirement of the synesthetic system and its software pertains to the choreographic aspect of synesthetic art and to the ability of the system to adapt its visual elements to interesting scenarios of change and movement within the event space. This time/dynamic component is clearly choreographic in character, whether actual dancers are projected into the event space, or whether the "dancers" consist of abstract shapes, colors, textures, or wisps of smokey ephemera moving about and through one another.

This choreography has three aspects. The first is a choreography of change associated with such terms as mutation, permutation, transformation, and metamorphosis; this has a topological aspect as well. The second is a choreography of movement in space, a shift from here to there, or of continuous movements over looping and interweaving paths of motion within the event space. (See the panel labeled "object motion and motion paths" in the block diagram.) And the third imposes a choreographic aspect on the timing of the change and motion events, which can accelerate and decelerate, and involve modulated shifts of timing as complex and subtle as the graceful movement of a ballerina, whose art consists as much in the timing of a movement as in the sculptural shape and arc of the movement itself. (See the panel labeled "time/change/motion synesthetics" in the block diagram.)

## Music and Sound

Just as the visual/spatial aspect of synesthetic art is able to draw upon and enlarge the entire body of resources of computer graphics, the musical/acoustic aspect is able to do the same within the closely associated fields of electronic and computer music. Important among these resources are the many customized interactive devices (keyboards, pedals, sliders, dials, the insertion and extraction of floppy disks) developed for composing and improvising computer music. Also important is the fact that sounds, like images, can be either captured from the real world by microphones, or synthesized through electronic or digital techniques (ref. 7).

Most important is the computer spatialization of sound that is so central to this concept of bimodal synesthetic art. Evidence is plentiful that the existing, well-proven technique of spatializing sounds by computer is growing in use and aesthetic effectiveness. For example, just within the past year, a spatialized composition of computer music was incorporated into a 45-ft open-form sculptural construction as a bimodal mix that is almost borderline synesthetic (ref. 8). In fact, if the sculpture itself, which is completely immobile, were kinetic in some way, and if the spatialized sounds interacted meaningfully with the kinetic aspect of the structure, the work might approach the synesthetic.

## Production and Performance

The system and its software must provide its users with the means to work in a variety of modes for creating many different kinds of synesthetic art. In addition to working directly and interactively with the system, the user should be able to take advantage of intelligent robotic and quasi-robotic support when it is needed for a specific purpose—i.e., a fast-moving improvisation in which the performer(s) could not possibly keep everything in hand without intelligent robotic support from the system. This robotic-type support is not only helpful, it is absolutely indispensable when the system is sustaining an ongoing "hands-off" performance, a special way of using the system that, depending on the inclination of the user, may involve either intermittent, frequent, or constant intervention into what the system is doing. (If the intervention is constant, the user has switched by definition into the fully interactive mode.)

Within these automated productions, an important subset is the transductive mode, yet another hands-off situation that essentially replaces the artist as sole intervenor, with intermittent or ongoing interventions from a variety of sources—interventions which are continuously mediated and structured by discriminating and aesthetically "sensitive" robotic components within the software. These intervening agencies in turn are made up of ambient energies, signals, and other "information" such as light, heat, sounds, vibrations, barometric pressure, brain waves, heartbeat, traffic patterns—many of the endless possibilities have for years been incorporated into diverse forms of environmental, transductive, and "systems" art, some of them computerized, most of them not (ref. 9). Clearly, synesthetic art produced in the transductive mode will acquire many specific forms for many different kinds of users and applications, all of them so readily interchangeable that the distinction between an amateur and a professional performance will tend to blur (or will, that is, if the artist working with the synesthetic system wants it that way).

Not least important is the fully robotic mode, in which the system, driven by a program that the artist has set up (or more rarely, may even have written, or possibly expanded), behaves like an

autonomous artist in its own right in producing either a continuous, ongoing work in the performance mode, or a series of individual productions in the serial-robotic mode. This fully robotic approach is not as far-fetched as it may seem; variations of it have been used for years by some of the pioneers of computer art in this country, Europe, and elsewhere (ref. 10).

## TWO PROJECTS ON THE FRINGE OF SYNESTHETIC ART

Synesthetic art as a concept has yet to produce an actual example of the genre to discuss or reproduce here. Nevertheless, I have been involved with a number of projects peripheral to synesthetic art over the years. These can be used to illuminate the subject, but should not be mis-construed as examples of what is still an art of the future. From these I have selected two projects, the first as an example of software oriented strictly to the robotic mode, and the second for its combination of both interactive and robotic possibilities.

### Applications of the Serial-Robotic Mode

An example of software capable of generating graphics in a serial-robotic mode is the largest and most complicated of the programs I have designed and developed to date. Called SHAPE3D, it was written during the middle 1970s with the help of two talented student programmers primarily as an experiment in the serial-robotic design of sculpture. In addition, however, the project reflects my long-standing conviction that the computer, in addition to its contribution to the creative aspect of art, also will foster a new approach to research in art theory and aesthetics. More specifically, the idea concerns a highly promising synergism of theory and practice between (1) the use of suc-cessive series of serial-robotic productions as an innovative and potentially powerful approach to computer-based research in art theory and the principles of design; and (2) the testing of the rules, principles, compositional devices, etc., generated by this research through their use in the serial-robotic production of various kinds of computer art. Of course, synesthetic art is obviously the kind of computer art with the most to gain from this valuable source of robotic intelligence con-cerning formal/syntactic structure-inducing algorithms and devices (refs. 11-14).

Operating with a vocabulary of 64 modular block-like elements and a set of 30 input parame-ters, SHAPE3D is capable of generating serial-robotic runs of as many as 50 or 100 or more graphics at a time, with never a duplicate composition in any series. The six graphics comprising the group reproduced as figure 2 were selected from a number of different serial-robotic runs to demonstrate the range of variations in style that can be obtained through various settings of the 30 parameters. The single unframed graphic in the group of six was selected from a serial-robotic run of 150 compositions, the best of which was chosen as a model for the complete sculpture shown as figure 3 (ref. 15).

### Calligraphic Stereo-Sculpture

For 3 months during the fall of 1978, I collaborated with an associate on a project that used a StereoRealist camera to record sequences of stereoscopic light calligraphies of the kind shown in figure 4. Inspired by a famous Gjon Mili strobe photograph showing Picasso drawing in space with a pen light, we purchased the stereo camera, collected an assortment of flashlights, colored

gels, luminous objects, and objects that could be illuminated (including a number of translucent plastic buckets) and set up a kind of event space in front of the stereo camera. As many as eight successive swoops and splashes in space were superimposed on the film in the camera to create each of about 30 stereo-calligraphies. A setup not unlike the one described here would be useful for collecting a large repertory of paths of motion on which to graft varieties of images and sounds. Or a variation could be used to capture the events *in toto*–the rich colors and textures along with the underlying paths. Or alternatively, the effects and the paths could be simulated through software, or through combinations of capture and mathematical synthesis (ref. 16).

## CONCLUSION

A concept of a new form of art called synesthetic art has been described, along with the characteristics of the computer-based system required for its production. The profoundly computer-oriented character of this form of art informs its relevance to themes and topics such as interactive graphics, virtual 3-D displays and projection systems, user-system ergonomics, artificial intelligence, robotics, and telerobotics. Preparing this paper has caused me to rethink, expand, and clarify my thinking on synesthetic art, and has left me even more convinced of its significance and virtual inevitability for the future of art. The progress of computer stereographics, in particular, makes it especially timely to begin thinking about actual start-up projects in stereo-synesthetics–not just a single project, but many of them, as the task is so multifarious and the directions that can be taken so diverse. In addition, this paper should assure those readers involved in fields related to spatial displays and instrumentation that aspects of NASA-sponsored research may have implications beyond NASA itself, beyond industry, business, and other obvious areas of possible application. For spatial displays are relevant to art, especially to that kind of art which is computer-based, time-variant, synesthetic, and looks forward to what is going to happen in the next century.

# REFERENCES AND NOTES

[1]The importance of the technical aspect of art is especially evident in the history of Western music and the evolution of those technological marvels, the instruments comprising the modern symphony orchestra. Likewise, the history of Western painting since the Renaissance would be unthinkable except for the invention of the oil medium and its enormous virtuosity and pliability in comparison to encaustic and tempera, whose stylistic possibilities are far more constrained.

[2]My concept of synesthetic art began in the early 1940s as a form of projected kinetic sculpto-painting with music. After 15 years on the shelf, I revived and expanded the idea in 1967 on realizing that the computer made some form of illusory-projective synesthetic art not only a feasible, but a virtually inevitable, development over the long term.

[3]The Atari-based stereo goggles can be purchased for less than $150 under the name of Stereo-Tek from Antic Publishers, Inc., 524 Second Street, San Francisco, CA 94104. Better computer stereo-graphic systems having higher resolutions are also on the market.

[4]Until recently, another approach to 3-D image generation by computer was available on the market for computer-aided design and other potential applications. Called SpaceGraph, a product of Genisco Computers Corporation, the system combined a graphic display with a small vibrating mirror to generate black-and-white images within a virtual display area measuring 20x25x30 cm. Its future now seems problematic.

[5]In addition to its lack of motion parallax, stereoscopy, in respect to the components of spatial perception, almost routinely violates the way in which they normally function in synchronous gestalt patterns. As for holography, though far superior to stereoscopy "in principle," it hardly bears comparison with stereoscopy in terms of practical computer applications and potential for real-time operation.

[6]The Turing test is the classic test for artificial intelligence proposed by Alan Turing, the British mathematician and computer scientist. Questions are passed to a computer and a human respondent hidden behind a curtain and the answers are passed back in written form. When it is impossible to distinguish between the answers provided by the computer and those by the human respondent, the computer can be said to have the level of intelligence of a human being. I proposed my "Turing test" for 3-D projection as a note at the end of my article "Computer Sculpture: Six Levels of Cybernetics," Artform, May 1969.

[7]J. Chowning, "The Simulation of Moving Sound Sources," JAES, Preprint no. 726(M-3) for the 38th convention, 1970.

[8]The work, a collaboration of sculptor Sherry Healy of Chicago and Charles Bestor, professor of electronic and computer music at the University of Massachusetts at Amherst, was first presented at the Chicago International Art Exposition at Navy Pier in May 1987. The sculpture consists of four open-form modular units made of wood that as a group add up to an impressive 47- by 12- by 9-ft installation. Four sets of four speakers embedded in portions of the sculpture provide the electronic and computer-generated music, which is recorded on tape for replay every 20 min. The sounds, as they execute varied choreographic patterns between four sets of speakers, are heard differently in different parts of the work by those circulating around and through it.

[9]In my article on computer sculpture (ref. 6), I also introduced the term "transductive art" as a generic expression covering all forms of kinetic and environmental art driven by some form of energy, information, or signal originating from outside the work or system itself.

[10]Pioneering work in the automated production of computer art goes all the way back to the early 1960s and to the farsighted experiments in "generative art" carried out by a group of German aestheticians, computer scientists, and artists. The most impressive contribution in this vein so far is that of Harold Cohen of University of California at San Diego, whose program Aaron, begun in 1972, continues to expand and become increasingly intelligent, autonomous, and powerful. As a virtual surrogate of Cohen's personality as an artist, it even succeeds in demonstrating "talent."

[11]Virtually all computer-based research in art is currently in stylistics, a field that traditionally has been focused on the exhaustive description and analysis of a particular style of painting, sculpture, or architecture. Under the impact of the computer, however, this information is beginning to be tested through incorporation into programs capable of generating visually credible simulations of the style undergoing study. Through devising shape grammars appropriate to the targeted style, this technique has been applied to the architecture of Palladio (ref. 12), to the work of the Russian non-objective painter Kandinsky (ref. 13), and to that of the American abstract painter Diebenkorn (ref. 14).

[12]G. Stiny and W. Mitchell, "The Palladian Grammar," Environ. and Plan. B, vol. 5, 1978.

[13]R. Lauzzana, "A Measurement of Image Concordance Using Replacement Rules," IEEE Conference on Systems, Man, and Cybernetics, Atlanta, 1986.

[14]J. Kirsch and R. Kirsch, "Computer Grammars for the Syntactical Analysis of Paintings," in Proc. of the 26th International Congress of the History of Art, Washington, D.C., 1986.

[15]This modular composition was created for an exhibition of the University of Massachusetts at Amherst sculpture faculty held in one of the university galleries during the spring of 1978. The work, which measured 12 by 17 ft at the base, consisted of a subset of the complete 64-block set that SHAPE3D is capable of generating, manipulating, and plotting as serial-robotic compositions. The modular blocks, enlarged to conform to the proportions of those used in the design generated by SHAPE3D, were constructed of Masonite and painted white.

[16]My associate, Michael Friedman, and I alternated between creating the swaths of color-in-space for superimposing on the film, and handling the camera, which involved dropping a black cloth over the open lens while the luminous "brush" was being replaced with another for the next calligraphic event.

# Block Diagram of the Synesthetic Supermedium

User Input and Control Modes

Real-World Image/Acoustic Capture

Computer-Generated Image/Acoustic Material

Transductive Inputs — Real-World/Artificial

## Visual/Spatial Components

| Rotation & Multiple Rotations | Objects | Object Attributes | Time/Change/Motion Synesthetics |
| | Arrays | Object Appearance/Disappearance | |
| | Regions | Object Transformations | |
| | Total Event Space | Object Motion & Motion Paths | |
| | Multiple Event Spaces | | |

Image/Sound Synesthetic Integration

Musical/Acoustic Components

Performance/Narrative/Literary Components

## Total Synesthetic Mix

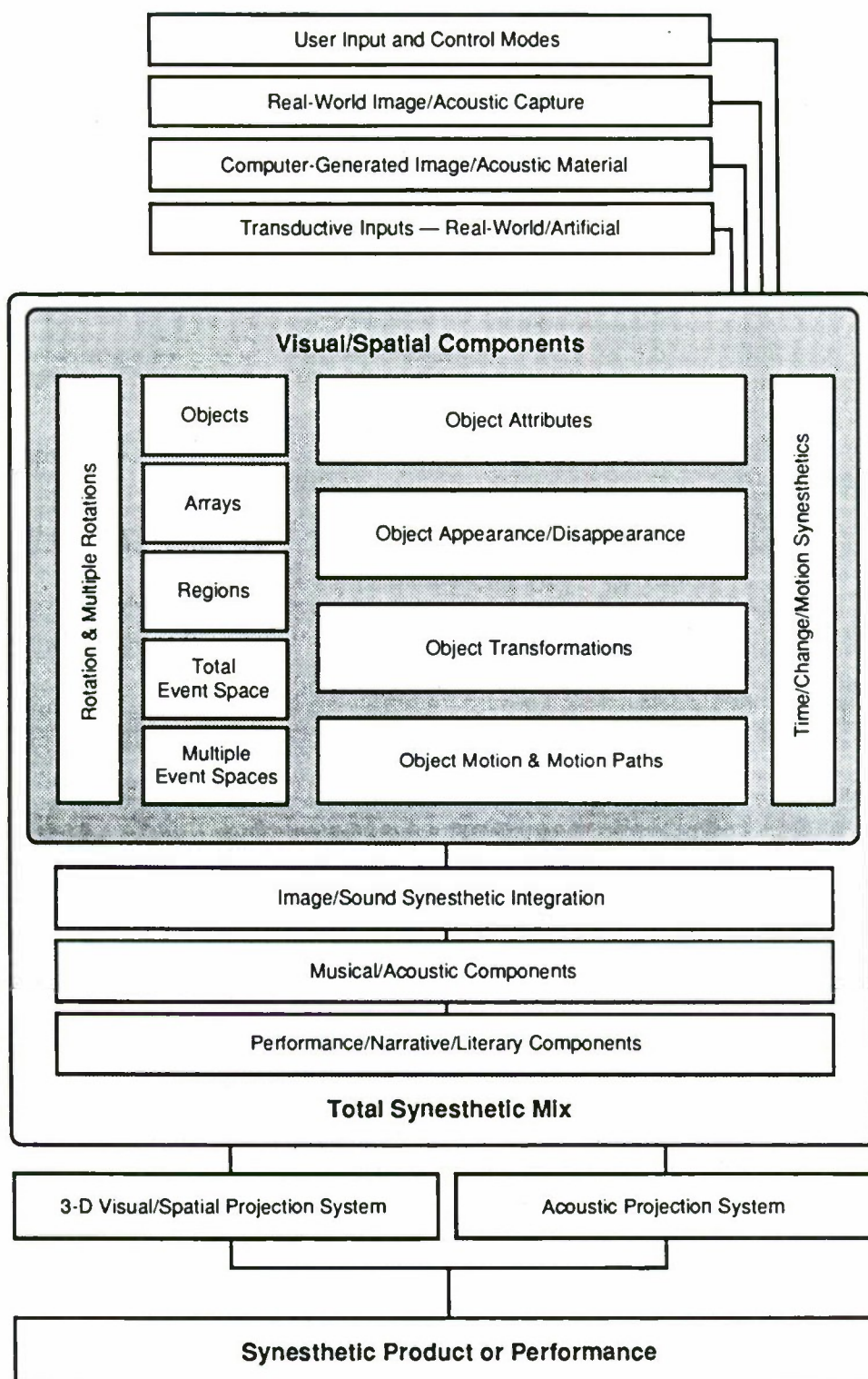| 3-D Visual/Spatial Projection System | Acoustic Projection System |

## Synesthetic Product or Performance

Figure 1.– Synesthetic supermedium.
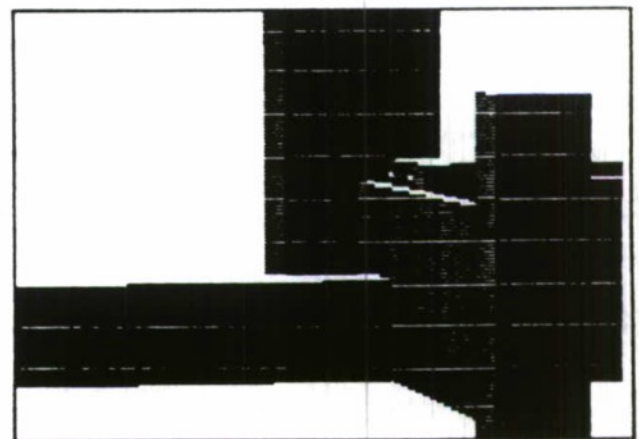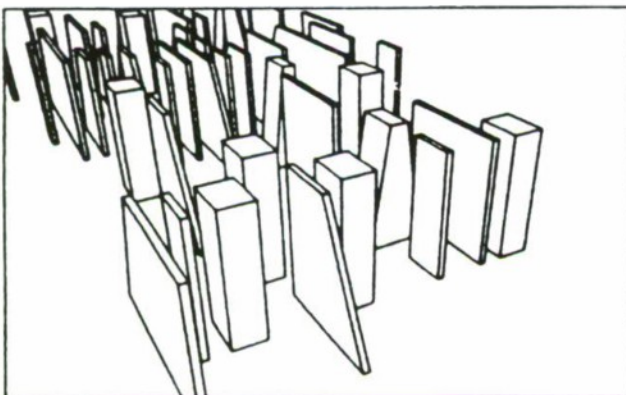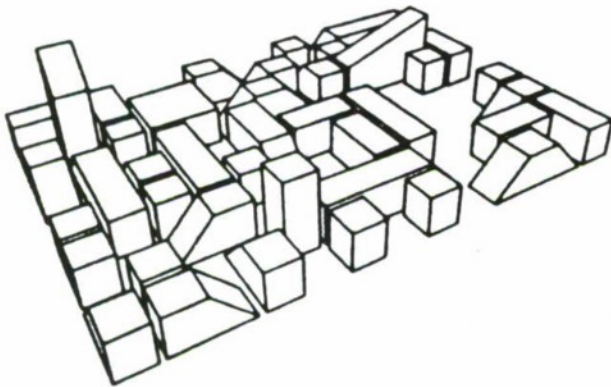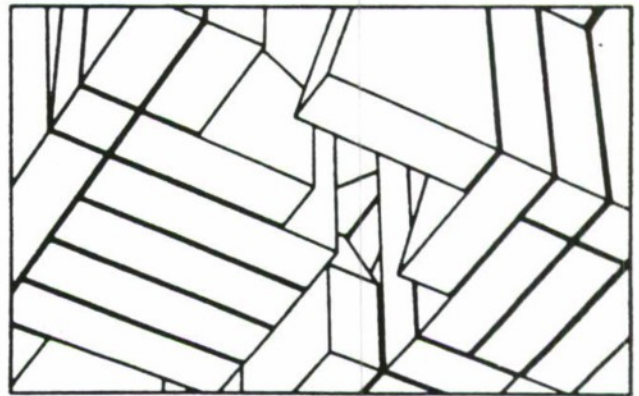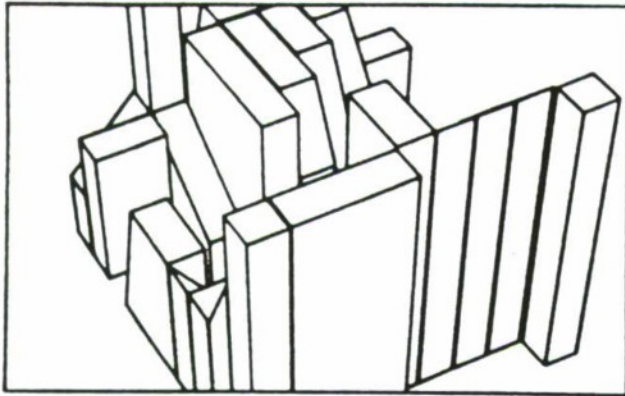
Figure 2.– Variations of serial-robotic runs.
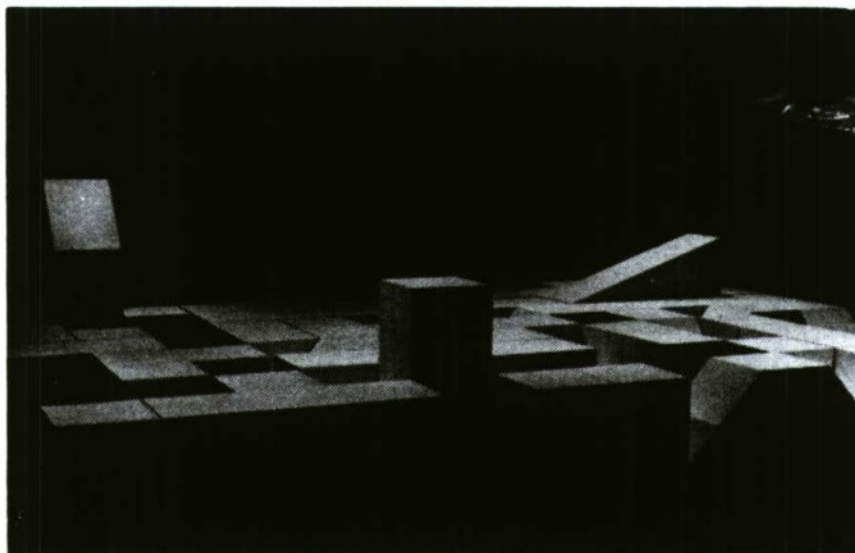
Figure 3.– Complete sculpture chosen from a serial-robotic run of 150 compositions.

Figure 4.– Stereoscopic light calligraphy example.

# WIDE-ANGLE DISPLAY DEVELOPMENTS BY COMPUTER GRAPHICS

William A. Fetter
Research Director, SIROCO
2165 156th Avenue Southeast
Bellevue, Washington

## SUMMARY

Computer graphics can now expand its new subset, wide-angle projection, to be as significant a generic capability as computer graphics itself. My purpose is to present you with some prior work in computer graphics leading to an attractive further subset of wide-angle projection, called hemispheric projection, to be a major communication media. Hemispheric film systems have long been present and such computer graphics systems are in use in simulators. This is the leading edge of capabilities which should ultimately be as ubiquitous as CRTs. The credentials I have for making these assertions are not from degrees in science or only from my degree in graphic design, but in a history of computer graphics innovations, laying groundwork by demonstration. I believe it is timely to look at several development strategies, since hemispheric projection is now at a point comparable to the early stages of computer graphics, requiring similar patterns of development again.

## POLARITY

Nobel Prize winner, Dr. Herbert Simon of Carnegie-Mellon University, in his book SCIENCE OF THE ARTIFICIAL, characterized the natural sciences as the pursuit of "what is," and the sciences of the artificial (which includes design), as the pursuit of "what should be." It occurs to me that NASA, more than any institution in history, has to stretch itself to the extreme ends of these polarities as well as cover the complete spectrum between. In designing vital systems, it must reach into the future, championing far-sighted objectives while using the most rigorous scientific knowledge, especially human performance. Each of these polarities has an organizational counterpart which can effect patterns of achievement. In the early stages of a new development, I believe it is fitting and effective to operate in the "what should be" mode, with attention to, and migration toward, the "what is" mode.

## BACKGROUND

Computer graphics efforts have included a number of research and development paths such as simulations of cockpit visibility, human figure performance, operations analysis and wide-angle projection. Many of these paths wer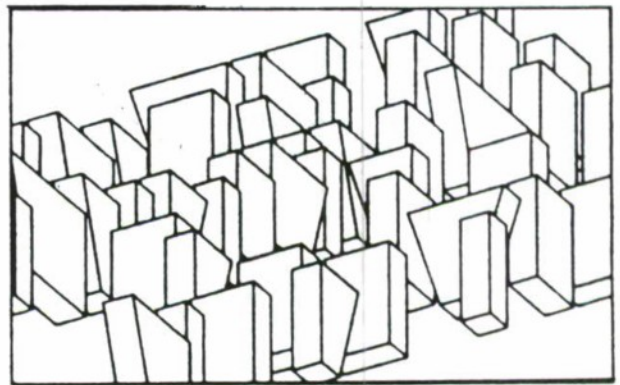e firsts and many of these were followed up over decades in three work environments, Boeing, SIU-C, and SIROCO. This work often stimulated others by showing "what to do," helping to spawn some of the computer graphics capabilities we see today.

The approaches taken can be usefully applied to the development of an array of hemispheric-projection display-system applications.

## COMPUTER GRAPHICS

The term computer graphics was coined about my initial work at the Boeing Company in 1959. I cannot claim that I coined the term, as some have suggested, because, in reality my supervisor, Verne Hudson, both authorized my proposal to work in this area and further suggested shortening my longer project title to just the two words.

This effort began with a research letter defining a near-term effort. It also listed the ultimately sought attributes of computer graphics, which included many of the visual characteristics in the field today. This work also achieved the landmark Bernhart-Fetter patent on perspective images generated by digital computer. An organization was assembled to form a close relationship between research, demonstrations and direct applications to needed tasks.

The overall goals of more accurate, reliable, and clear images are sought in advancing hemi-spheric display systems.

The precursor to my computer graphics innovations at Boeing was a hand plot, which I then illustrated in the process of designing a book. During graphic design assignments at the University of Illinois Press Art Division, I designed the book SPACE MEDICINE for Werner Von Braun. I felt that an illustration of his space station concept should appear in orbit on the title page and that it should be as accurate as possible, in part, an homage to Chesely Bonestele. So that it would be precise, I plotted points by hand, using a technique that eliminated the vanishing points then taught in schools. The tiresome degree of repetition in the process and the emerging claims for computer capabilities convinced me that at some time in the future I would have a computer assist this process.

Now let us look at the efforts at the Boeing Company during the 1960s by glimpsing several lines of research and applications to aerospace requirements.

1.    Eye: All of our activity was directed to more effectively reach the eye/brain complex in support of engineering design.

2.    Computer Interior: The task was to utilize any existing computer system available to us at Boeing in order to carry out the production of useful images and series of images.

3.    Communication Need: We developed an approach of defining our communication work within a spectrum of needs to be met.

4.    Communication Media: We made every effort to relate the need to specific media and to integrate computer graphics into that flow.

5.    Boeing 747: Static output was produced using computer graphics axonometrics and per-spectives such as this Boeing 747. We merged our work with such related capabilities as master dimensions.

6. Carrier landing: More dynamic applications included dozens of color/sound motion pictures, on all major Boeing designs of the 1960s.

7. SST Mockup: Support to mockups included the Supersonic Transport 60-ft-wide diorama of precise views at the 100-ft decision point.

8. First Man: Human figure simulations were applied to 747 and space cockpit studies of reach and instrument vision, using 100 body sizes.

9. Hemispheric: Preliminary software was demonstrated for stimulus material to be projected on the interior surface of a hemisphere.

10. Interactive: Studies of interactive human factors computer graphics included anthropometrics, visibility, and other applications.

(Our disseminations stimulated other manufacturers' work. For example, GE, seeing our Runway Visual Range studies, was able, with their outstanding capability, to produce more advanced fog simulations.)

11. 747 Polar Plot: An early purpose for wide-angle projections, in this case a Mercator projection, was the first computer graphics polar plot to aid in meeting FAA requirements for the Boeing 747 visibility.

12. Screen Angle: Our efforts to explore wider viewing angles made it desirable to gain further human factors information such as Dreyfuss.

13. Human Factors in Design: In seeking out information we wanted to design systems not interfering with other human factors parameters.

14. Pacific Science Center: This hemispheric display facility for films designed by Boeing in Seattle was useful and convenient.

15. First Test: Some of the early tests did not yield a perfect match and the geometry of the software had to be rewritten.

16. Room Test: The next successful tests included one showing visual effects of sitting in a square room viewed inside a hemisphere.

17. 747 Cockpit: Among the test applications made was the 747 cockpit windows displayed as seen from the interior.

(We also proposed to use the hemispherics in the E series 747 aircraft for high-level decision makers to rapidly apprehend complex displays.)

18. Vulnerability: An application to vulnerability studies used the similarity of hemispheric geometry to the geometry of airburst threats.

19.   NASA:  A potential application with NASA Public Relations was to use telemetered displays for better public understanding of the space effort, including output to television or hemispheric facilities.

Now let us look at our hemispheric path of work at Southern Illinois University at Carbondale during the 1970s, to apply this to more comprehensive design issues.

1.   Computer Graphics Research:  At the SIU-C Department of Design in the 1970s, we conducted further computer graphics research under the sponsorship of the SIU-C Research and Projects Office, the National Science Foundation, and other private sector sources.

2.   Association of Science/Technology Centers:  As an outgrowth of the earlier NASA public relations study and the new goals at SIU to develop Buckminster Fuller's advanced concept of a World Resources Simulation Center, we again looked at the potential of existing hemispheric facilities that could convey necessary information to the public.  A related project involved an SIU committee on Earth Resources and a period of time spent at NASA to determine types of satellite imagery available that might be processed through this type of facility.

3.   Pacific Science Center Spacearium:  During the 1970s, the modest research funding levels limited the tests to projecting glass slides.  Mose of these centers have geometry which does not exactly match.

4.   70-MM Wide-Angle Film:  Sample film from the Spacearium shows the identical distortions our test plots matched.  Members of the Psychology Department at SIU-C found the possibilities for group interaction and decision-making in such a system to be promising.  Among the more obvious advantages were the wide field of view, absence of extraneous visual elements, and the resulting complete attention by the observer.  Among the more obvious disadvantages were the cost, complexity, and size of the systems then available or fundable to build.

Now let us look at the hemispheric research path at SIROCO, an independent research institute, in the 1980s.

1.   Yards, Feet, Inches:  At SIROCO, the perimeter folding problem was solved and special attributes of hemispheric displays were studies.  One attribute was maintaining orientation within a display of a hierarchy of facts and images.  A simple example here is yards, feet, and inches.  While the full effect cannot be seen in a simple flat slide example, the advantages are more apparent in hemispheric images.

2.   Earth:  To demonstrate the value of a capacity for great changes in scale, needed for a world resources center, a long zoom was created.

3.   United States:  The zoom continues toward the United States.

4.   Illinois:  We continue, showing Illinois county boundaries.

5.   Carbondale:  And on to the street grid of Carbondale, Illinois.

6.   Human Figures:  And finally to the scale of two human figures.

7.    Color: The images can be in color.  The sequence outlined previously ends with our human figure, which is based on only one data base of an infinite number of accurate surface definitions of anthropometric percentiles and somatotypes.  This rendering was done in a joint activity with the Lawrence Livermore National Laboratory using Frank Crow's HILITE, Steve Williams' updates.

In 1978 at SIROCO, we made a proposal to NASA on hemispheric display.  This was approved for scientific merit; however, it could not be funded.  In 1981, in assisting the SIG-GRAPH committee which sponsored the annual meeting in Seattle, we worked successfully to reinstate the showing of Nelson Max's IMAX film demonstration of wide angle.  In 1984 our original work helped stimulate SIGGRAPH's OMNIMAX film production.

Where is hemispheric going?  I believe the answer is EVERYWHERE.  At NASA, both hemispheric and spheric displays are already used in existing and emerging simulators.  In future space flights, hemispheric projection should find its way into the crew's flight deck, work stations, and entertainment stations.  In communicating with computers, there is just as large a bottleneck at the visual interface as at the internal bottlenecks that gave rise to parallel processing.  Hemispheric projection can contribute solutions.  Elsewhere, hemispheric technologies that emerge should benefit from economy-of-means in both computing and visual systems.  Only a small proportion of a complete hemispheric image needs to be generated for many applications using head-mounted displays.  With the costs for computer capacities dropping dramatically, even processing all the pixels should become practical for more applications.

## CONCLUSION

There are fundamental human factors issues involved in this new tool.  We should build generic systems rather than reinvent each application.  We should, I believe, develop a location for multipurpose breadboard demonstrations with the balanced support and stimulus of a wide variety of relevant technological expertise.  Further, we should explore whole new communication modalities such as "Orientation Graphics," "Discovery Graphics," and "Analogy Graphics."  Spin-offs in miniatured, low-cost systems should find their way into offices and work stations.

I have presented my personal experiences over a period of years because there are elements of these early holistic approaches needed now.  NASA may be the best institution in which to explore this since at NASA, as in hemispheric displays, we are at just the beginning of practical visions of the future that are all about us.

# VOLUMETRIC VISUALIZATION OF 3D DATA

Gregory Russell[1] and Richard Miles
Department of Mechanical and Aerospace Engineering
Princeton University
Princeton, New Jersey

## INTRODUCTION

In recent years, there has been a rapid growth in the ability to obtain detailed data on large complex structures in three dimensions. This development occurred first in the medical field, with CAT scans and now magnetic resonance imaging, and in seismological exploration. With the advances in supercomputing and computational fluid dynamics, and in experimental techniques in fluid dynamics, there is now the ability to produce similar large data fields representing 3D structures and phenomena in these disciplines.

These developments have produced a situation in which currently we have access to data which is too complex to be understood using the tools available for data reduction and presentation. Researchers in these areas are becoming limited by their ability to visualize and comprehend the 3D systems they are measuring and simulating.

## HISTORY

In response to this, there is growing activity in the area of visualization of 3D data. Some early work in this area was done by Harris et al. (1979) at the Mayo Clinic and Herman et al. (1984) at the University of Pennsylvania in the area of medical imaging. In 1983, Jaffey, Dutta, and Hesselink (1984) approached the subject from a different direction. They developed the "source-attenuation" model, and used holograms to visualize 3D subjects. More recently, there is stronger emphasis on interactive visualization, and concentration on techniques and systems for general use and commercial products (Goldwasser, 1985; Hunter, 1984).

Much of the recent activity is directed toward improving and extending the use of graphics techniques for interactive visualization of data based on surface representations. The groundwork for this was done by Herman et al. Work in this area is continuing both in academic groups (Herman at the University of Pennsylvania (Herman et al., 1984 and Fuchs at North Carolina (Fuchs et al., 1985), and in several commercial ventures (notably CEMAX)). Also, graphics projects at NASA, JPL, and aerospace corporations have been providing increasing support for visualization tasks based on conventional graphics concepts.

The more interesting projects involve departures from conventional graphics. By careful use of transparency, it is possible to produce images of 3D systems which provide true volumetric visualization, rather than surface projections. We have been working on this type of system for the past three years (Russell and Miles, 1987), concentrating on techniques which are efficient enough

---

[1]G. Russell currently at IBM T. J. Watson Research Lab, Yorktown Heights, N.Y. 10598.

to be used interactively on existing computer systems. Pixar Corporation has recently been developing a package to support volumetric visualization, including an approach called Volume Rendering Technique, which they developed with Phillips Medical Systems and Dr. E. Fishman (1987) of Johns Hopkins University. This package is perhaps the most comprehensive image-based system commercially available at this time.

An approximation to volumetric imaging is also provided in PLOT3D, a graphics software system developed at JPL. This package includes a facility for producing nested transparent contour surfaces from a volumetric data base, which provides surprisingly good visualization of the data. Its primary limitations are data size (about 100,000 data points) and the number of contours it can support. Also, since this is a rather symbolic representation, it must be interpreted with care.

## VOLUMETRIC VS. 2 1/2D VISUALIZATION

Normal pictorial illustration (stills), and most widely used 3D graphics techniques are limited to providing 2 1/2D surface images. That is to say, along any line of sight there is only one object or surface visible. This usually produces pictures from which a rough idea of the three-dimensional structure of the original scene can be deduced. In contrast, X-ray images generally do not have a unique interpretation as projections of some three-dimensional subject, and even X-ray stereo pairs are insufficient to provide an unambiguous interpretation without a priori knowledge about the subject.

This is a computational constraint which applies not only to visual observation of pictures, but to interpretation of volumetric projections in general. Vision, however, is capable of limited volumetric perception and comprehension, if given adequate stimulus.

In order to achieve effective volumetric perception, it is necessary to present volumetric data in a form that vision is accustomed to dealing with. While cross sections are often useful for detailed study of internal features, it is difficult or impossible to fully comprehend the 3D structure of an object in this manner. Instead, data must be presented as we would see a real object. Natural visual processing transforms this information back into a mental structural model. Volumetric characteristics of the data are conveyed by making the projection TRANSPARENT, as implied in the earlier discussion.

The requirements for volumetric perception are basically the same as for computed axial tomography. A set of projection images from many different viewpoints is computationally sufficient to reconstruct the internal details of a subject. Visual reconstruction has several added constraints: the images must be presented as an ordered sequence of closely spaced views, and they must be shown at a rate of at least 8 to 10 frames/sec. These constraints are dictated by the temporal character of visual perception.

For perception of volumetric structure (rather than surface structure), complex optical phenomena such as lighting and shading, specular (surface) reflections, and diffraction and diffusion are not useful. In fact, these effects generally make the basic structure of volumetric scenes more difficult to understand, overwhelming the viewer with fine details and optical distortions. Simple luminance and opacity are adequate for volumetric visualization.

# SYSTEM IMPLEMENTATION

We have developed a system at Princeton which implements this approach to volumetric visualization on a PC/AT (Russell and Miles, 1987). The algorithms upon which it is based are efficient enough to provide a usable off line visualization system on the AT (precomputed images take approximately 1 min/view for 2 million data points) and they are suitable for development into a real-time interactive visualization system using current state-of-the-art commercial hardware (AT&T Pixel Machine, for example).

The model for the system has the following characteristics.

1. Data consists of samples on any regular 3D lattice (e.g., simple cubic, face-centered cubic, hexagonal close packed).

2. The data elements are treated as nebulous, fuzzy regions localized around the sample coordinates. (i.e., no subvoxel definition–consistent with proper sampling technique).

3. Optical model includes luminance and opacity control at each data point, with the possibility of handling a light source (no refraction or specular reflection).

4. Views are computed directly from the data, without any intermediate representation. This reduces the risk of artifacts and avoids simplification of the data that may lead to the loss of features.

5. Perspective is not supported (this is subordinate to motion).

This combination of characteristics yields a model which is well-behaved and computationally efficient, with enough flexibility to provide a broad range of visual effects.

The implementation on the PC/AT operates in a two-step process. For a given data base, a sequence of views is computed, based on a selected set of optical characteristics onto which the data are mapped, and a viewpoint and axis of rotation for the data. Each image takes about 60 to 75 sec, for a typical data base of 2 million samples (e.g., 32x256x256 or 128x128x128), and we usually generate anywhere from 15 images (for a restricted range of views) to 120 images (for a full rotation of the data). The images are stored on a disk as they are generated. When a sequence is complete, the images are loaded by a second program for viewing. Up to 180 clipped images (176x176) may be loaded into 6 Mbytes of RAM on the PC/AT. They may then be viewed as a movie on a full-color, 8-bit greyscale display at frame rates up to 15 frames/sec. The viewpoint is controlled interactively using a mouse, within the precomputed range.

# EVALUATION

This method of visualization provides good comprehension for a range of subjects and optical characteristics. Its most significant advantage is that it is very robust. There is little or no preprocessing of the data, so there are generally no computational artifacts. Even data containing no distinct surfaces can be accurately visualized, since this method does not rely on surfaces as the

fundamental elements of a scene. The use of motion as the means of communicating structure allows all the data to be made visible through the use of transparency. This provides a high degree of confidence in the resulting visualization. It is also robust in the sense that an informative set of images can be produced using simple optical characteristics (luminance = data value, high transparency) with little or no a priori knowledge about the data itself.

The motion/transparency approach is most effective with scenes of moderate complexity (such as that shown in Fig. 1), that is, scenes whose structure can be largely comprehended as a whole. With very complex scenes, containing perhaps hundreds of detailed components (e.g., a video cassette recorder guts), this type of visualization suffers from showing too much information, which cannot be fully comprehended as a single entity.

## COMPLEXITY

The issue of complexity arises in visualization for two distinct reasons. The first is the visual limitation just mentioned. The mind is incapable of performing a complete internal reconstruction of a volumetric scene, as is done in a CAT scan, for example. We have observed that beyond a certain level of complexity in depth (apparently three to four layers of structure), the mind's ability to maintain a conceptual model of a scene begins to fail.

In addition to the visual/conceptual limitation, there is an optical constraint which limits the degree of complexity which is practically acceptable. There is a tradeoff between the amount of transparency used (which affects the visibility of embedded structures) and the amount of contrast available in small features. This is directly related to signal-to-noise (S/N) ratio. Vision does not have particularly large S/N ratio, so fine details quickly lose definition as transparency is increased. This is also a limiting factor in CAT scans, but the devices used have much higher S/N ratios, so much lower contrast can be tolerated in CAT-scan source images than is detectable visually.

These considerations provide strong motivation to develop means of reducing and controlling the level of complexity in volumetric visualization.

## THE ROLE OF BINOCULAR VISION

From a very early point in our investigation of visualization, it was clear that stereo pairs were inadequate as illustration of volumetric scenes. Once we had a working visualization system based on motion, it was easy to see how much more comprehensive this approach is than static stereo viewing. For some time, we assumed that adding stereopsis to the motion-based system would not be worthwhile, since static experiments suggested that stereopsis would not work well on precisely those scenes where some improvement was needed. Specifically, scenes with extensive volumetric content and high complexity, such as medical data, generally have low contrast and few clearly defined, unique features on which stereopsis can operate. For scenes which are visualized with low transparency, which provides more distinct features, stereopsis is not really needed since these scenes are generally quite easily understood with only the motion-based visualization.

When we actually were able to try out stereo and motion together, the results were somewhat surprising. With scenes of medical data with moderate to high transparency, static stereo viewing is relatively ineffective, as expected. However, when motion and stereo viewing are used together, the stereopsis provides noticeable enhancement to the visual perception of the structure over motion alone. There is apparently some interaction between the visual mechanisms which use stereo and motion to deduce structure. The combined effectiveness suggests that stereopsis is facilitated by information made available by motion, which perhaps allows better feature matching between images, resulting in more and better disparity measurements.

This strong interaction between stereopsis and motion perception means that stereopsis must be considered as an important part of any visualization system. Though motion is very powerful alone, considerable enhancement is possible through the use of binocular vision.

## CONCLUSIONS

This approach to visualization, using transparency and motion in an image-based system, has significant advantages over systems based on solid rendering or graphical modeling. Most significant are the broader range of volumetric structure which can be visually represented and the robustness and freedom from artifact which volumetric visualization provides. A comprehensive visualization facility should certainly include the ability to perform both image-based and graphical rendering, and in the future these techniques should be increasingly integrated to allow both graphical and image-based components in a single visualization.

Computers are now becoming available which will be capable of performing visualization tasks interactively. This will dramatically change the way in which visualization is used, particularly for very complex subjects. As interactive visualization becomes more practical, the current emphasis on development of techniques for data reduction and rendering should be supplanted by the need for means of controlling and interacting with the visualization process. As the potential degrees of freedom for controlling a visualization increase with the complexity and size of scenes, the design of effective control mechanisms will be a difficult endeavor.

Some simple control mechanisms, such as clipping, spatial editing tools, and 3D cursors, are relatively easy to implement. However, for complex data, control mechanisms should parallel the way in which structures are decomposed and manipulated conceptually. This means providing the capability to specify the structural components of a scene and control their visual characteristics by referring to them as objects. Automated or computer-aided object segmentation is required to make this practical, but for the purpose of interactive control of visualizations, the accuracy and reliability of segmentations need not be as high as it must for conventional, noninteractive visualization.

Additionally, it may be useful to be able to produce geometric distortions of data in order to push obstructing objects out of the way without separating them altogether from the region of interest. The net effect would be to produce the equivalent of an exploded view for structures of nondiscrete components. This would be particularly useful in medical applications. If information about connectivity and stiffness can be incorporated into the process, this could make the visualization system even more useful in surgical training or preoperative planning environments, where the mechanical properties of tissue structures is very important.

Advanced modes of interaction will become more and more important as volumetric display is applied to more ambitious problems of data interpretation.

# REFERENCES

E. Fishman et al., "Volumetric Rendering Techniques: Applications for Three Dimensional Imaging of the Hip," Radiology 163, No. 3, 737 (1987).

H. Fuchs et al., "Fast Spheres, Shadows, Textures, Transparencies, and Image Enhancements in Pixel Planes," ACM Computer Graphics Vol. 19, No. 3, 111 (1985).

S. Goldwasser et al., "Physician's Workstation with Real-Time Performance," IEEE Comput. Graphics Appl. Vol. 5, No. 12, CGA.00, 44 (1985).

L.D. Harris, R.A. Robb, T.S. Yuen, and E.L.Ritman, "Display and Visualization of Three-Dimensional Reconstructed Anatomic Morphology: Experience with the Thorax, Heart, and Coronary Vasculature of Dogs," J. Comput. Assist. Tomogr. 3, 439 (1979).

G.T. Herman, R.A. Reynolds, and J.K.Udupa, "Computer Techniques for the Representation of Three-Dimensional Data on a Two Dimensional Display," Proc. Soc. Photo-Opt. Instrum. Eng. 507, 3 (1984).

G. Hunter, "3D Frame Buffers for Interactive Analysis of 3D Data," Proc. Soc. Photo-Opt. Instrum. Eng. 507, 178 (1984).

S. Jaffey, K. Dutta, and L. Hesselink, "Digital Reconstruction Methods for Three-Dimensional Image Visualization," Proc. Soc. Photo. Opt. Instrum. Eng. 507, 155 (1984).

G. Russell and R. Miles, "Display and Perception of 3-D Space-filling Data," Appl. Optics, Vol. 26, No. 6, 973 (1987).

Figure 1.– Vortex rings resulting from the Crow instability. Navier-Stokes simulation data provided by Dr. Micheal Shelley, Princeton University.

# FINAL PROGRAM

## Spatial Displays and Spatial Instruments:
## A Symposium and Workshop
## Sponsored by the
## National Aeronautics and Space Administration
## and the
## University of California, Berkeley

**August 31 – September 3, 1987**
**Asilomar, California**

## August 31

| | |
|---|---|
| 2:00–5:00 pm | Check-in and Orientation |
| 4:00–5:00 pm | Reception |
| 6:00 pm | Dinner |

7:00–8:00 pm     Welcomes
D. Nagel
Chief: Aerospace Human Factors
Ames Research Center

Conference Purpose
   Pictorial Communication
   S. R. Ellis
   Ames Research Center

   The Role of Pictorial Communication in Aerospace
   M. W. McGreevy
   Ames Research Center

Conference Logistics, etc.
   M. Moultray and S. R. Ellis

## September 1

7:30–8:20 am     Breakfast

8:20–12:30 pm     Invited Paper Session 1
Chairman: S. R. Ellis

## SPATIAL PERCEPTION

8:30 am        "Perspectives on Perspective"
Professor R. L. Gregory
University of Bristol Medical School
Introduction by S. R. Ellis

5-min discussion

9:05 am        "Visual Realism in Boeing Simulators"
C. Kraft
Formerly Boeing Commercial Aircraft Company
Introduction by J. Cutting

5-min discussion

9:40 am        10-min Coffee Break

## SPATIAL ORIENTATION

9:50 am        "Perception of Egocentric Visual Direction"
Professor I. Howard
York University
Introduction by H. Mittelstaedt

5-min discussion

10:25 am        "Egocentric Direction in Simulators"
T. Furness
Wright-Patterson Air Force Base
Introduction by S. Fisher

5-min discussion

## PICTURE PERCEPTION

11:00 am        "Picture Perception and Virtual Space"
Professor H. A. Sedgwick
SUNY College of Optometry
Introduction by J. Perrone

5-min discussion

11:35 am        "The Design of Pictorial Displays"
Professor S. Roscoe
New Mexico State University
Introduction by J. Hartzell

5-min discussion

| 12:20–1:30 pm | Lunch Break |
| 1:30–5:50 pm | Contributed Paper Session<br>Chairman: M. Kaiser |

## SPATIAL PERCEPTION

| 1:30 pm | "Spatial Factors Influencing Stereopsis and Fusion"<br>Professor C. Schor<br>U.C. Berkeley |
| 1:50 pm | "Scaling Stereoscopic Space"<br>Professor J. Foley<br>U.C. Santa Barbara |
| 2:10 pm | "Paradoxical Monocular Stereopsis and Perspective Vergence"<br>Professor J. T. Enright<br>Scripps Institution of Oceanography |
| 2:30 pm | "The Perception of Three Dimensionality Across Continuous Surfaces"<br>Professor K. Stevens<br>University of Oregon |
| 2:50 pm | "Perceiving Environmental Properties From Motion Information: Minimal Conditions"<br>Professor D. Proffitt<br>University of Virginia<br>and<br>M. Kaiser<br>Ames Research Center |

## SPATIAL ORIENTATION

| 3:10 pm | "Memory Distoritons of Visual Displays"<br>Professor B. Tversky<br>Stanford University |
| 3:30 pm | 20-min Coffee Break |

## PICTURE PERCEPTION

3:50 pm "The Effect of Changes in Viewpoint on the Pictorial Perceptions of Spatial Layout and Orientation Relative to the Observer"
Professor B. Goldstein
University of Pittsburgh

4:10 pm "Cinematic Efficacy, or What the Visual System Did Not Evolve to Do"
Professor J. Cutting
Cornell University

4:30 pm "Congruence Under Motion as a Basis for the Perceived Geometrical Structure of Forms andSpaces"
Professor J. Lappin
Vanderbilt University
and
Dr. T. Wason
ALLOTECH

4:50 pm "A Theoretical Analysis of the Recognition of Pictorial Displays"
Professor I. Biederman
SUNY Buffalo

5:10 pm "Spatial Displays and Spatial Instruments from the Graphics Design Perspective"
A. Marcus
Aaron Marcus Associates

5:30 pm "Interactive Displays in Medical Art"
Professor D. McConathy
University of Illinois

6:00–7:30 pm Dinner

8:00–10:00 pm        Poster Sessions and Informal Discussion

"Synthetic Perspective Optical Flow: Influence on
Pilot Control Tasks"
T. Bennett, W. Johnson, and J. Perrone
Ames Research Center
and
A. Phatak
Analytical Mechanics Associates
Ames Research Center

"Visual Enhancements and Control in Telerobotics"
W. S. Kim, F. Tendrick, and Professor L. Stark
U.C. Berkeley

"Visual Slant Underestimation"
J. Perrone and P. Wenderoth
Ames Research Center and University of Sydney

"Optical and Gravitational Information in the
Perception of Eye Level"
Professor A. Stoper and M. Cohen
Ames Research Center

"Interactive Spatial Instruments for Proximity
Operations" (video)
Professor A. Grunwald and S. R. Ellis
Ames Research Center

"Exocentric Direction Judgements Based on Pictorial
and Real-World Layouts" (video)
S. R. Ellis, Professor A. Grunwald, and S. Smith
Ames Research Center

"Criteria for the Successful Representation of
Information"
Professor M. Hagen
Boston University

"Development of a Stereo 3-D Pictorial Primary
Flight Display" (video)
M. Nataupsky
Langley Research Center
and
T. Turner, H. Lane, and L. Crittenden
Research Triangle Institute

"Representational Structure for Evaluation of Human/Robotic System Control"
K. Corker
BBN Laboratories Incorporated

"Adaptation to Non-Zero Disarrangement of the Visual Field"
Professor R. Welch and M. Cohen
Ames Research Center

"Theoretical Issues in the Development of a 2-D and 3-D Computer-Aided Designer Support System"
J. Hartzell
Ames Research Center

"Telepresence in Dataspace" (video)
S. Fisher
Ames Research Center

"The Photo-Colorimetric Space as a Medium for the Representation of Spatial Data"
K. F. Kraiss and H. Widdel
Forschungsinstitut für Anthropotechnik

"The Role of Attensity in Spatial Perception"
M. Companion
Lockheed-Georgia Company

"Achieving a Concrete 'UP': Embodiment of Spatial Relationships in a Head-Mounted Display System" (video)
W. Robinett
Ames Research Center

"Requirements and Features of a Synesthetic Supermedium"
Professor R. Mallar
ATARI Computer
Arstechnica: Center for Art and Technology
University of Massachussetts

"Helmet Mounted Displays—Spatial Orientation Problems"
S. Hart
Ames Research Center

"How to Reinforce Perception of Depth in Single
Two-Dimensional Pictures"
S. Nagata
NHK Science and Technical Research Laboratory

"Direction of Movement Effects Under Transformed
Visual-Motor Mappings"
H. Cunningham and Professor M. Pavel
Stanford University

"Efficiency of Graphical Perception" (video)
Y. Gu, Professor G. Legge, and A. Luebker
University of Minnesota

"Applications of Human Factors for Cartography and
Geography"
Professor George F. McCleary
University of Kansas

"Interactive Digital Video Interface to an Atlas of
Histology"
Michael D. Doyle
University of Illinois, Urbana-Champaign

## September 2

| | |
|---|---|
| 7:30–8:30 am | Breakfast |
| 8:30–12:00 am | Invited Paper Session 2<br>Chairman: S. R. Ellis |

### MANIPULATIVE CONTROL

| | |
|---|---|
| 8:35 am | "Visuo-Motor Plasticity and Time Lags"<br>R. Held and N. Durlach<br>MIT<br>Introduction by D. Fadden<br><br>5-min discussion |
| 9:10 am | "Displays and Controls for the Space Shuttle Arm"<br>G. M. McKinnon<br>CAE Electronics Ltd.<br>Introduction by B. Bridgeman<br><br>5-min discussion |
| 9:45 am | 10-min Coffee Break |

## VESTIBULAR ASPECTS

9:55 am          "Theories of Visual-Vestibular Interaction"
C. Oman
MIT
Introduction by R. Haines

5-min discussion

10:40 am          "Vestibular Realism in Simulators"
J. Sinacori
Consulting engineer
Carmel, California
Introduction by E. Palmer

5-min discussion

## COMPUTER GRAPHICS

11:15 am          "Graphics Hardware and Software: Coming
Attractions"
F. Baskett
Silicon Graphics Inc.
Introduction by (to be determined)

5-min discussion

11:50 am          "The Making of the Mechanical Universe"
J. F. Blinn
JPL Graphics Laboratory
Introduction by M. Kaiser

5-min discussion

12:30–1:45 pm      Luncheon
Speaker: J. P. Allen
Space Industries Inc.
(former Shuttle astronaut)
"The Challenges of Flying the Manned Maneuvering Unit in Earth Orbit"

1:50-5:30 pm      Contributed Papers
Chairman: A. Grunwald

## MANIPULATIVE CONTROL

1:50 pm          "Two Modes of Visual Representation"
Professor B. Bridgeman
U.C. Santa Cruz

| 2:10 pm | "Perception-Action Relationships Reconsidered"<br>Professor W. Shebilske<br>Texas A&M University |
|---|---|
| 2:30 pm | "A Computer Graphics System for Visualizing<br>Spacecraft in Orbit"<br>D. Eyles<br>Charles Draper Laboratories |
| 2:50 pm | "Displays for Telemanipulations"<br>B. Hannaford, M. Salganicoff, and A. Bejczy<br>Jet Propulsion Laboratory |
| 3:10 pm | "Experience in Teleoperation of Land Vehicles"<br>D. McGovern<br>Sandia National Laboratories |
| 3:30 pm | "Spatial Displays and Pilot Control: Where Do We<br>Go From Here?<br>D. Fadden, R. Braune, and J. Wiedemann<br>Boeing Commercial Airplane Company |
| 3:50 pm | 20 min Coffee Break |

## VESTIBULAR ASPECTS

| 4:10 pm | "Determinants of Space Perception in<br>Weightlessness"<br>Professor H. Mittelstaedt<br>Max Planck Institut für Verhaltensphysiologie |
|---|---|
| 4:30 pm | "Voluntary Presetting of the Vestibular Ocular Reflex<br>Permits Gaze Stabilization Despite Perturbation of<br>Fast Head Movements"<br>Professor W. Zangemeister<br>Neurologische Klinik der Universität<br>Hamburg |

## COMPUTER GRAPHICS

| | |
|---|---|
| 4:50 pm | "Wide Angle Display Developments by Computer Graphics"<br>W. A. Fetter<br>Siroco |
| 5:10 pm | "Visualizing Space Filling Data"<br>G. Russell<br>Princeton University |
| 6:00–8:00 pm | BBQ on Asilomar Terrace |

### September 3

| | |
|---|---|
| 7:30–8:30 am | Breakfast |
| 8:30–9:00 am | Summary Session<br>Summary: L. Stark/U.C. Berkeley<br>Thanks to all<br>Checkout |
| 10:00 am | Leave for tours of Ames Research Center |
| 11:30 am | Arrive Ames Research Center |
| 11:30–12:30 pm | Lunch at Ames Cafeteria (not included in conference fee) |
| 12:30–2:30 pm | Open House at Aerospace Human Factors Division and possibly Vestibular Research Facility |
| 2:35 pm | Leave for Return to Asilomar |
| 4:00–4:15 pm | Arrive Monterey Airport/Asilomar Conference Center |

# RESEARCH CONTACTS

Terry Abbott
NASA Langley Research Center
MS 152e
Hampton, VA 23665
Phone: (804) 865-3917

\* Joseph P. Allen
Executive Vice President
Space Industries
711 W. Bay Area Blvd.
Webster, TX 77598
Phone: (713) 338-2676

Herbert Archer
Lockheed Georgia Co.
Department 72-75
Marietta, GA 30063
Phone: (404) 424-4713

\* Forest Baskett
Silicon Graphics
2011 Steirlin Rd.
Mountain View, CA 94043
Phone: (415) 960-1980

\* Professor Richard Beindorf
1855 Folsom St.
Suite 521
San Francisco, CA 94103
Phone: (415) 476-4468

Antal K. Bejczy
Jet Propulsion Laboratory
MS 198-330
4800 Oak Grove Dr.
Pasadena, CA 91103
Phone: (818) 354-4568

\* Tom Bennett
NASA Ames Research Center
MS 239-3
Moffett Field, CA 94035
Phone: (415) 694-5960

Rainer Bernotat
Forschungsinstitut für Anthropotechnik
D-5307 Wachberg-Werthoven
Bundes Republik Deutschland

Professor Irving Biederman
Department of Psychology, Park Hall
SUNY Buffalo
Amherst, NY 14260
Phone: (716) 636-3650/3682/3671

\* James F. Blinn
Jet Propulsion Laboratory
MS 510-113
4800 Oak Grove Dr.
Pasadena, CA 91109
Phone: (818) 397-9699/9051

\* Professor Bruce Bridgeman
Department of Psychology
U.C. Santa Cruz
Santa Cruz, CA 95064
Phone: (408) 429-4005

Anthony Busquets
NASA Langley Research Center
MS 152e
Hampton, VA 23665
Phone: (804) 865-3917

+ Sherry Chappell
NASA Ames Research Center
MS 239-3
Moffett Field, CA 94035
Phone: (415) 694-6909

Dr. Susan Chipman
Office of Naval Research
1142CS
800 N. Quincy St.
Arlington, VA 22217-5000
Phone: (202) 696-4318

\*Contributed Paper
+Attended Conference

William S. Cleveland
AT&T Bell Laboratories
Rm 2C276
600 Mount
Murray Hill, NJ 07974
Phone: (201) 582-6861

* Malcolm M. Cohen
NASA Ames Research Center
MS 239-7
Moffett Field, CA 94035
Phone: (415) 694-6441

Professor H. Steven Colburn
Department of Engineering
Boston University
100 Cummington St.
Boston, MA 02215
Phone: (617) 353-4342

Clay Coler
NASA Ames Research Center
MS 239-3
Moffett Field, CA 94035
Phone: (415) 694-5716

* Michael A. Companion
Lockheed Georgia Co.
Department 72-23, Zn419
Marietta, GA 30063
Phone: (404) 424-3819/4395

* Kevin Corker
Bolt, Berenek, & Neuman Co.
10 Moulton St.
Cambridge, MA 02238
Phone: (617) 491-1850/3065

* Helen Cunningham
Psychology Department
Jordan Hall, Building 420
Stanford University
Stanford, CA 94305
Phone: (415) 725-2432

* Professor James Cutting
Psychology Department
Uris Hall
Cornell University
Ithaca, NY 14853-7601
Phone: (607) 255-2000/6305

+ Professor Diana Damos
Department of Human Factors
I-SSM
University of Southern California
Los Angeles, CA 90089-0021
Phone: (213) 548-0399

+ Theodore Demosthenes
ALPA
1149 Snowberry Ct.
Sunnyvale, CA 94087
Phone: (408) 735-1712
       (213) 413-4530 msgs

* Michael Doyle
University of Illinois, College of Medicine
Room 190, Medical Sciences Building
506 S. Mathews
Urbana, IL 61801
Phone: (217) 333-9627

* Nathaniel I. Durlach
Research Laboratory of Electronics
Building 36-709
MIT
Cambridge, MA 02139
Phone: (617) 253-2534

William Edmunds
Engineering & Air Safety Department
ALPA
1625 Massachusetts Ave., NW
Washington, DC 20036
Phone: (703) 689-4198

Dean Jay Enoch
School of Optometry
Minor Hall
U.C. Berkeley
Berkeley, CA 94720
Phone: (415) 642-3414

*Contributed Paper
+Attended Conference

*   Professor James T. Enright
    Scripps Institution of Oceanography
    U.C. San Diego
    La Jolla, CA 92093
    Phone: (619) 534-3784

*   Don Eyles
    Charles Stark Draper Laboratory
    555 Technology Square
    Cambridge, MA 02139
    Phone: (617) 258-2460

*   Delmar M. Fadden
    Boeing Commercial Airplane Company
    P. O. Box 3707
    MS 77-70
    Seattle, WA 98124-2207
    Phone: (206) 237-0173

    Dr. Pierre Falzon
    Groupe de Psycho. Ergonomique
    INRIA, Domaine de Voluceau
    Rocquencourt, BP 105
    78153 Le Chesnay Cedex, France
    Phone: (1) 39 63 55 11  (3) 954 90 20

    James Farber
    AT&T Bell Labs
    H01J-333
    Crawfords Corner Rd.
    Holmdel, NJ 07733
    Phone: (201) 949-4412

*   W. A. Fetter
    Siroco
    2165 156st Ave. SE
    Bellevue, WA 98007
    Phone: (206) 746-9512

*   Scott Fisher
    NASA Ames Research Center
    MS 239-3
    Moffett Field, CA 94035
    Phone: (415) 694-6789

*   Professor John M. Foley
    Department of Psychology
    U.C. Santa Barbara
    Santa Barbara, CA 93106
    Phone: (805) 961-2030/2791

*   Thomas Furness
    AFHRL Wright-Patterson AFB
    4070 Hyland Dr.
    Dayton, OH 45424
    Phone: (513) 255-7601/7602

    Steven Gano
    1120A Montgomery St.
    San Francisco, CA 94133
    Phone: (415) 421-3611

    Gabriel Gautier
    Universite de Provence
    Rue Poincare
    13397 Marseille CEDEX 13, France

    Professor Barbara Gillam
    Department of Psychology
    University of New South Wales
    Canberra, Australia

    Professor Walter Gogel
    Department of Psychology
    U.C. Santa Barbara
    Santa Barbara, CA 93106
    Phone: (805) 961-2045

*   Professor Bruce Goldstein
    Department of Psychology
    University of Pittsburgh
    Pittsburgh, PA 15213
    Phone: (412) 624-4533

*   Julian Gomez
    NASA Ames Research Center
    RIACS, MS 230-5
    Moffett Field, CA 94035
    Phone: (415) 694-6142

*Contributed Paper

*   Professor Richard L. Gregory
    Department of Psychology
    University of Bristol
    Bristol England BS8 1TD
    Phone: UK 44 272 303 030   272 739 701

    Dr. Michael Griffin
    ISVR
    University of Southhampton
    Southhampton, England S09 5NH
    Phone: (703) 559-122

*   Professor Arthur J. Grunwald
    Department of Aeronautical Engineering
    TECHNION
    Haifa, Israel
    Phone: 972 4 292302

*   Yuanchao Gu
    Department of Psychology
    University of Minnesota
    75 E. River Rd.
    Minneapolis, MN  55455
    Phone: (612) 625-0846/4042

    Ralph Haber
    Department of Psychology
    University of Illinois
    Chicago Circle
    Chicago, IL  60680
    Phone: (312) 996-5263

*   Professor Margret A. Hagen
    Department of Psychology
    Boston University
    64 Cummington St.
    Boston, MA  02215
    Phone: (617) 353-2075

+   Richard Haines
    NASA Ames Research Center
    RIACS MS 230-5
    Moffett Field, CA  94035

*   Blake Hannaford
    Jet Propulsion Laboratory
    MS 198-330
    4800 Oak Grove Dr.
    Pasadena, CA 91103
    Phone: (818) 354-0351

*   Sandra Hart
    NASA Ames Research Center
    MS 239-3
    Moffett Field, CA  94035
    Phone: (415) 694-6072

*   James Hartzell
    NASA Ames Research Center
    MS 239-3
    Moffett Field, CA  94035
    Phone: (415) 694-5743

    Jack Hatfield
    NASA Langley Research Center
    MS 152e
    Hampton, VA  23665
    Phone: (804) 865-3917

    Dr. Harold Hawkins
    ONR
    1142CS
    800 N. Quincy St.
    Arlington, VA  22217-5000
    Phone: (202) 696-4323

    Thomas Heckman
    Department of Psychology
    York University
    4700 Keele St.
    North York, Ontario, Canada M3J 1P3
    Phone: ((416) 736-2100x2441

*   Professor Richard Held
    MIT
    Department of Brain and Cognitive Science
    Cambridge, MA  02139
    Phone: (617) 253-1000/5745

*Contributed Paper
+Attended Conference

Robert Hennesy
NASA Ames Research Center
LHX
Moffett Field, CA 94035
Phone: (415) 694-4024

Professor Julian Hochberg
Psychology Department
Columbia University
Schermerhorm Hall, Room 406
New York City, NY 10027
Phone: (212) 280-1754/3683

Captain Houghton
NAMRL
NAS
Pensacola, FL 32508
Phone: (904) 452-3286

* Professor Ian P. Howard
Department of Psychology
York University
4700 Keele St.
North York, Ontario, Canada M3J 1P3
Phone: ((416) 736-2100x2441

Joyce Iavechia
Engineering Psychology
Naval Air Development Center
Code 6022
Warminster, PA 18974
Phone: (215) 441-1499

Professor George Jenks
Department of Geography
University of Kansas
Lawrence, KS 66044
Phone: (913) 864-5143/4775

* Walter W. Johnson
NASA Ames Research Center
MS 239-3
Moffett Field, CA 94035
Phone: (415) 694-6249

+ Kevin Jordan
Psychology Department
San Jose State University
San Jose, CA 95292
Phone: (408) 277-2793

* Mary Kaiser
NASA Ames Research Center
MS 239-3
Moffett Field, CA 94035
Phone: (415) 694-6243

* Won Soo Kim
Bldg. 278
Jet Propulsion Laboratory
4800 Oak Grove Dr.
Pasadena, CA 91103
Phone: (818) 354-5047

Professor Stephen Kosslyn
Harvard University
1236 William James Hall
33 Kirkland St.
Cambridge, MA 02138
Phone: (617) 495-1000

William Kovaks
Wavefront Technology
530 E. Montecito
Santa Barbara, CA 93103
Phone: (805) 962-8117

* Conrad L. Kraft
12210 S.E. Tenth
Bellevue, WA 98005
Phone: (206) 747-2896 (home)

* Karl F. Kraiss
Forschungsinstitut für Anthropotechnik
D-5307 Wachberg-Werthhoven,
Bundes Republik Deutschland
Phone: Telefon (0228) 852-1
Telex 885589 fgan-d

*Contributed Paper
+Attended Conference

* Ronald Kruk
  CAE
  Box 1800
  Saint-Laurent
  Quebec, Canada H4L4X4
  Phone: (514) 341-6780

Jeff Kulick
MIT-Arch. Machine Group
225 Grove St.
Newton, MA 02166
Phone: (617) 253-5114/5960

Robert Langridge
Computer Graphics Laboratory
926-Medical Sciences
U.C. San Francisco
San Francisco, CA 94143-0446
Phone: (415) 476-2630

* Professor Joseph Lappin
  Psychology Department
  Vanderbilt University
  Nashville, TN 37240
  Phone: (615) 322-2398
         (615) 373-8985

James Larimer
NASA Ames Research Center
MS 239-3
Moffett Field, CA 94035
Phone: (415) 694-5150

+ Terri Lawton
  Jet Propulsion Laboratory
  MS 23
  4800 Oak Grove Dr.
  Pasadena, CA 91109
  Phone: (818) 354-4257/6508

George Lee
U.S. Geological Survey
345 Middlefield Road
MS 531
Menlo Park, CA 94025
Phone: (415) 323-8111

* Professor Gordon Legge
  Department of Psychology
  University of Minnesota
  75 E. River Rd.
  Minneapolis, MN 55455
  Phone: (612) 625-0846/4042

Professor Herschel W. Leibowitz
Department of Psychology
Pennsylvania State University
160 Sandy Ridge Rd.
State College, PA 16803
Phone: (814) 237-6398

Dr. Alan M. Lesgold
Learning Research & Development Center
Bldg. 516
University of Pittsburgh
Pittsburgh, PA 15260
Phone: (412) 624-4141

Frank Lewendouski
CAE-Link
1077 Arques Ave.
Sunnyvale, CA 94088-3484
Phone: (408) 720-5500

Andrew L. Lippay
CAE
Box 1800
Saint-Laurent
Quebec, Canada H4L4X4
Phone: (514) 341-6780

Professor Jack Loomis
Department of Psychology
U.C. Santa Barbara
Santa Barbara, CA 93106
Phone: (805) 961-2475 Dept. 2791

* Professor Robert Mallary
  Department of Art
  University of Massachusetts
  Amherst, MA 01003
  Phone: (413) 545-1902/2630

*Contributed Paper
+Attended Conference

*   Aaron Marcus
    Aaron Marcus & Associates
    1196 Euclid Ave.
    Berkeley, CA  94708-1640
    Phone: (415) 527-6224/6225

*   Professor George McCleary
    Department of Geography
    University of Kansas
    Lawrence, KS  66045-2121
    Phone: (913) 864-5143

*   Professor Deirdre McConathy
    Department of Biocommunication Arts
    University of Illinois
    1919 West Taylor
    Chicago, IL  60612
    Phone: (312) 996-8340

*   D. McGovern
    Advanced Technology Division
    Sandia National Laboratory
    Albuquerque, NM  87185
    Phone: (505) 844-1542

*   Michael W. McGreevy
    NASA Ames Research Center
    MS 239-3
    Moffett Field, CA  94035
    Phone: (415) 694-6170

    Donald J. Meagher
    OCTREE Corp.
    4404 Adragna Court
    San Jose, CA  95136
    Phone: (408) 972-9086

*   Horst Mittelstaedt
    Max Plank Institute
    für Verhaltensphysiologie
    8131 Seewiesen
    Federal Republic of Germany
    Phone: (49) 81.57.291

    Mel Montemerlo
    Automation and Robotics
    NASA Headquarters
    Code RC
    Washington, DC  20546
    Phone: (202) 453-2787

+   Jeffrey Mulligan
    NASA Ames Research Center
    MS 239-3
    Moffett Field, CA  94035
    Phone: (415) 694-5150

*   Shojirou Nagata
    NHK Science & Technical Research
     Laboratory
    10-11, kinuta 1-chome, Setagaya
    157 Tokyo, Japan
    Phone: 81 03 415 5111

+   Dave Nagel
    Apple Computer
    Cupertino, CA

*   Mark Nataupsky
    NASA Langley Research Center
    MS 152e
    Hampton, VA  23665
    Phone: (804) 865-3917

    Captain Joseph Oliver
    5097 Riverview Rd., N.W.
    Atlanta, GA  30327
    Phone: (404) 952-5859

*   Charles Oman
    Man-Vehicle Laboratory
    Rm 37-215
    Massachusetts Institute of Technology
    Cambridge, MA  02139
    Phone: (617) 253-7508

+   Everett A. Palmer III
    NASA Ames Research Center
    MS 239-3
    Moffett Field, CA  94035
    Phone: (415) 694-6073

*Contributed Paper
+Attended Conference

\* Misha Pavel
Department of Psychology
Jordan Hall, Build. 420
Stanford University
Stanford, CA 94305
Phone: (415) 725-2430

+ Ross Pepper
Ocean Systems Division
U.S. Naval Ocean Systems Center
P.O. Box 997
Kailua, HI 96734
Phone: (808) 254-4409

David Perkins
Project Zero/Harvard University
13 Appian Way
Cambridge, MA 02138
Phone: (617) 495-4343

\* John Perrone
NASA Ames Research Center
MS 239-3
Moffett Field, CA 94035
Phone: (415) 694-5150

\* Anil Phatak
NASA Ames Research Center
MS 239-3
Moffett Field, CA 94035
Phone: (408) 738-3650

Mark D. Phillips
Computer Technologies Associates
5670 S. Syracuse Circle #200
Englewood, CO 80111
Phone: (303) 889-1200

Professor Stephen Pinker
MIT E10-018
Department of Brain and Cognitive Science
Cambridge, MA 02139
Phone: (617) 253-8946

Steve Poltrock
MCC
8200 Neely Dr.
Austin, TX 78759
Phone: (512) 343-0978

\* Professor Dennis R. Proffitt
Department of Psychology
University of Virginia
Charlottsville, VA 22901
Phone: (804) 924-2251

Peter R. Rach
Air Line Transport Association of America
1709 New York Ave. NW
Washington, DC 20006
Phone: (202) 626-4023/31 (secretary)

John Reising
U.S. Air Force WAL
Flight Dynamics Laboratory
AFWAL/FIGR
Wright Patterson AFB, OH 45433
Phone: (513) 255-8274

Roger Remington
NASA Ames Research Center
MS 239-3
Moffett Field, CA 94035
Phone: (415) 694-5975

Professor James Richardson
Laboratoire de Physiologie du Mouvement
Universite de Paris-Sud
91405 Orsay Cedex, France
Phone: 01 588 3564 06 941 7073

\* Warren Robinett
NASA Ames Research Center
MS 239-1
Moffett Field, CA 94035
Phone: (415) 694-6433

Professor Irvin Rock
Psychology Department
U. C. Berkeley
Berkeyley, CA 94720

\* Stanley N. Roscoe

*Contributed Paper
+Attended Conference

Illiana Aviation Sciences Limited
Box 5095
Las Cruces, NM 88003
Phone: (505) 646-1408

R. Rosinski
HO1J-322
AT&T Bell Labs
Crawfords Corner Rd.
Holmdel, NY 07733
Phone: (201) 949-7633

Mary Ann Rudisel
Lyndon B. Johnson Space Center
SP341
2101 NASA Road 1
Houston, TX 77058
Phone: (713) 483-3706

+  Sylvie Rueff
Jet Propulsion Laboratory
MS 510-110
4800 Oak Grove Dr.
Pasadena, CA 91109
Phone: (818) 397-9695/9051

*  Gregory Russell
IBM Watson Research Lab

Dr. Dominique Scapin
Groupe de Psycho. Ergonomique
INRIA, Domaine de Voluceau
Rocquencourt, BP 105
78153 Le Chesnay Cedex, France
Phone: 1 776 0334    3 954 9021

*  Professor Clifton Schor
School of Optometry
419 Minor Hall
U.C. Berkeley
Berkely, CA 94720
Phone: (415) 642-1130

*  Professor H. A. Sedgwick
SUNY College of Optometry
100 E 24 St.
New York City, NY 10010
Phone: (212) 420-4900/5144

George Sexton
Lockheed-Georgia Co.
Department 72-23 Zone 419
Marietta, GA 30063
Phone: (404) 424-2184/2031

+  Michael Shafto
NASA Ames Research Center
MS 239-3
Moffett Field, CA 94035
Phone: (415) 694-6170

*  Professor Wayne L. Shebilske
Department of Psychology
Texas A&M
College Station, TX 77834
Phone: (409) 696-3444

Professor Thomas B. Sheridan
Department of Engineering
  and Applied Psychology
Massachusetts Institute of Technology
77 Massachusetts Ave. #3-346
Cambridge, MA 02139
Phone: (617) 253-2228/2201
       (617) 244-4181

+  Peter Shugart
USATRAC
Attn: ATRC-WS
White Sands Missile Range,
  NM 88002-5502
Phone: (505)

*  John Sinacori
John B. Sinacori Associates
Simulation & Arts Sciences
P. O. Box 360
Pebble Beach, CA 93953
Phone: (408) 373-1687

*Contributed Paper
+Attended Conference

Alvie Ray Smith
PIXAR
P. O. Box 13719
San Rafael, CA 94913
Phone: (415) 499-3600

Edward Hugh Spain
U. S. Naval Ocean Systems Center
P. O. Box 997
Kailua, HI 96734-0997
Phone: (808) 254-4434

Professor Kathryn Spoehr
Department of Psychology
Brown University
Providence, RI 02912
Phone: (401) 863-1000

* Professor Lawrence Stark
School of Optometry
481 Minor Hall
U. C. Berkeley
Berkeley, CA 94720
Phone: (415) 642-7196

* Professor Kent Stevens
Department of Computer Science
University of Oregon
Eugene, OR 97403
Phone: (503) 686-4408/4430

+ Lee Stone
NASA Ames Research Center
MS 239-3
Moffett Field, CA 94035
Phone: (415) 694-5150

R. J. Stone
British Aerospace
Naval Electronics Division
P. O. Box 5 Filton
Bristol BS12 7QW England
Phone: 272 693-831

* Professor Arnold Stoper
Department of Psychology
California State University
Hayward, CA 94542
Phone: (415) 881-3484/3469

Ivan Sutherland
Sutherland, Sproul and Associates
P. O. Box 1160
Palo Alto, CA 94302
Phone: (415) 725-2400

Professor Wayne Thiebaud
U.C. Davis Art Department
1617 7th Ave.
Sacramento, CA 95818
Phone: (916) 447-4980
          (916) 752-0105 (Art Dept.)

Professor W. R. Tobler
Department of Geography
U.C. Santa Barbara
Santa Barbara, CA 93106
Phone: (805) 961-3663/3831

David Tomko
NASA Ames Research Center
MS 242-3
Moffett Field, CA 94035
Phone: (415) 694-5723

Professor Edward R. Tufte
Yale University
1161 Sperry Rd.
Cheshire, CT 06410
Phone: (203) 432-5249

Dr. Michael T. Turvey
Haskins Laboratories
270 Crown St.
New Haven, CT 06510
Phone: (203) 865-6163

*Contributed Paper
+Attended Conference

* Professor Barbara Tversky
Department of Psychology
Jordon Hall, Building 420
Stanford University
Stanford, CA 94305
Phone: (415) 725-2400

Professor Andries Van Dam
Department of Computer Science
Brown University
Providence, RI 02912
Phone: (401) 863-1000/2340

+ Michael Venturino
AAMRL/HEA
Wright Patterson AFB
Dayton, OH 45433-6573
Phone: (513) 255-8895

Frank Vinz
Marshall Space Flight Center
Bldg. 4487
Huntsville, AL 35812
Phone: (205) 544-2121

Professor Hans Wallach
Department of Psychology
Swarthmore College
Swarthmore, PA 19081
Phone: (215) 328-8000x8436

* Thomas Wason
Allotech
715 W. Johnson St.
Raleigh, NC 27603
Phone: (919) 828-9446

+ Andrew Watson
NASA Ames Research Center
MS 239-3
Moffett Field, CA 94035
Phone: (415) 694-5419

Professor Daniel J. Weintraub
Human Performance Center
Perry Bldg.
330 Packard Rd.
Ann Arbor, MI 48104
Phone: (313) 763-0588

* Robert B. Welch
NASA Ames Research Center
MS 239-7
Moffett Field, CA 94035
Phone: (415) 694-5575

+ Beth Wenzel
NASA Ames Research Center
MS 239-3
Moffett Field, CA 94035
Phone: (415) 694-5716

+ Kendrick Williams
Douglas Aircraft Company
Dept. E-21, MS 35-36
3855 Lakewood Blvd.
Long Beach, CA 90846
Phone: (213) 513-2030/2020

+ Lawrence Wolpert
507 S. High St.
Yellow Springs, OH 45387
Phone: (513) 255-2894

Dr. Dan Wolz
ASHRL/MOE
Brooks AFB
San Antonio, TX 78235
Phone: (512) 536-1110

Lawrence Young
Man-Vehicle Laboratory
Room 37-215
Massachusetts Institute of Technology
Cambridge, MA 02139
Phone: (617) 253-7745

*Contributed Paper
+Attended Conference

\* Wolfgang Zangemeister
   Neurologische Klinik
   der Università Martinstr. 52
   D-2000 Hamburg
   West Germany
   Phone: (49) 40 460-1164

| NASA National Aeronautics and Space Administration | Report Documentation Page | | |
|---|---|---|---|
| **1. Report No.** NASA CP-10032 | **2. Government Accession No.** | **3. Recipient's Catalog No.** | |
| **4. Title and Subtitle** Spatial Displays and Spatial Instruments | | **5. Report Date** July 1989 | |
| | | **6. Performing Organization Code** | |
| **7. Editors** Stephen R. Ellis and Mary K. Kaiser and Arthur Grunwald (Technion, Haifa, Israel) | | **8. Performing Organization Report No.** A-88090 | |
| | | **10. Work Unit No.** 505-47-11 | |
| **9. Performing Organization Name and Address** Ames Research Center Moffett Field, CA 94035 | | | |
| | | **11. Contract or Grant No.** | |
| **12. Sponsoring Agency Name and Address** National Aeronautics and Space Administration Washington, DC 20546-0001 | | **13. Type of Report and Period Covered** Conference Publication | |
| | | **14. Sponsoring Agency Code** | |

**15. Supplementary Notes**

Point of Contact: Stephen R. Ellis, Ames Research Center, MS 239-3, Moffett Field, CA 94035 (415) 694-6147 or FTS 464-6147

**16. Abstract**

This conference proceedings brings together papers from two groups of researchers who ordinarily do not interact with each other, disciplinary scientists from academic and government laboratories and those engineers currently engaged in the design and testing of new visual displays incorporating sophisticated computer graphics. Others interested in pictorial communications have also contributed papers. These include several prominent computer scientists interested in graphics, several artists, medical illustrators, photogrammetrists, and geographers.

The proceedings consists of papers from invited lecture sessions. Brief critiques of the lectures are published as are papers based on the delivered lecturers.

The conference topics are divided into two main areas: (1) issues of spatial and picture perception raised by graphical electronic displays of spatial information and (2) design questions raised by the practical experience of designers actually defining new spatial instruments for use in new aircraft and spacecraft. Each topic is considered from both a theoretical and an applied direction. Emphasis is placed discussion of phenomena and determination of design principles.

| **17. Key Words (Suggested by Author(s))** Spatial perception, Pictorial communication Manual control, Visual-vestibular interaction Head-mounted displays | **18. Distribution Statement** Unclassified-Unlimited Subject Category – 54 | | |
|---|---|---|---|
| **19. Security Classif. (of this report)** Unclassified | **20. Security Classif. (of this page)** Unclassified | **21. No. of pages** 681 | **22. Price** A25 |

50 - 13